Rozprawa doktorska pt.

Optymalizacja technologii CRISPR-Cas9 w kontekście celowania w sekwencje powtarzające się

(ang. Optimization of CRISPR-Cas9 technology in the context of targeting repetitive sequences)

Magdalena Dąbrowska

Promotor:

dr hab. Marta Olejniczak, prof. ICHB PAN

Zakład Inżynierii Genomowej

Instytut Chemii Bioorganicznej PAN

Poznań, 2022

Składam serdeczne podziękowania

dr hab. Marcie Olejniczak
za lata wspólnej pracy, przekazaną wiedzę i doświadczenie

Koleżankom i Kolegom
z Zakładu Inżynierii Genomowej, Zakładu Genetyki Molekularnej
oraz Zakładu Biotechnologii Medycznej

Dziękuję Rodzinie i Najbliższym
za wyrozumiałość
i wsparcie każdego dnia

SPIS TREŚCI

FINANSOWANIE

Niniejsza praca powstała przy wsparciu finansowym udzielonym przez:

1. Narodowe Centrum Nauki
    - Grant Sonata Bis (2015/18/E/NZ2/00678)
    - Grant Etiuda 8 (2020/36/T/NZ2/00127)
2. Stypendium L'Oreal dla Kobiet i Nauki

LISTA PUBLIKACJI

Publikacje autorki, które wchodzą w skład pracy doktorskiej:

1. **Dabrowska, M.**, Juzwa, W., Krzyzosiak, W. J., & Olejniczak, M. (2018). Precise excision of the CAG tract from the huntingtin gene by Cas9 nickases. *Frontiers in Neuroscience*, *12*, 75. IF_{2020} 4,677
2. **Dabrowska, M.**, Czubak, K., Juzwa, W., Krzyzosiak, W. J., Olejniczak, M., & Kozlowski, P. (2018). qEva-CRISPR: a method for quantitative evaluation of CRISPR/Cas-mediated genome editing in target and off-target sites. *Nucleic Acids Research*, *46*(17), e101-e101. IF_{2020} 16,971
3. **Dabrowska, M.**, & Olejniczak, M. (2020). Gene therapy for Huntington's disease using targeted endonucleases. In *Trinucleotide Repeats* (pp. 269-284). Humana, New York, NY.
4. **Dabrowska, M.**, Ciolak, A., Kozlowska, E., Fiszer, A., & Olejniczak, M. (2020). Generation of new isogenic models of Huntington's disease using CRISPR-Cas9 technology. *International Journal of Molecular Sciences*, *21*(5), 1854. IF_{2020} 5,923

Inne prace autorki:

1. Ciesiolka, A., Stroynowska-Czerwinska, A., Joachimiak, P., Ciolak, A., Kozlowska, E., Michalak, M., **Dabrowska M.**, Olejniczak M., Raczynska K.D., Zielinska D., Wozna-Wysocka M., Krzyzosiak W.J., Fiszer, A. (2021). Artificial miRNAs targeting CAG repeat expansion in ORFs cause rapid deadenylation and translation inhibition of mutant transcripts. *Cellular and Molecular Life Sciences*, *78*(4), 1577-1596. IF_{2020} 9.261
2. Sledzinski, P., **Dabrowska, M.**, Nowaczyk, M., & Olejniczak, M. (2021). Paving the way towards precise and safe CRISPR genome editing. *Biotechnology Advances*, 107737. IF_{2020} 14.227

STRESZCZENIE

Choroby poliglutaminowe (poliQ) są neurodegeneracyjnymi chorobami genetycznymi spowodowanymi mutacją, która polega na zwiększeniu liczby powtórzeń trójki nukleotydów CAG w genach odpowiedzialnych za te choroby. Choroba Huntingtona (HD) występuje najczęściej ze wszystkich zaburzeń poliQ. Pomimo, że gen *HTT* odpowiedzialny za tę chorobę został dobrze scharakteryzowany, nadal nie są znane wszystkie aspekty jej patogenezy, co utrudnia stworzenie potencjalnej terapii.

Technologia CRISPR-Cas9 stwarza ogromne możliwości, zarówno na polu tworzenia lepszych modeli badawczych, jak również opracowania nowych podejść terapeutycznych. Wykorzystanie systemu CRISPR-Cas9 do badania sekwencji powtarzających się stanowi jednak duże wyzwanie technologiczne. Wydłużona sekwencja CAG tworzy drugorzędowe struktury RNA i DNA oraz przyczynia się do poślizgów polimerazy podczas amplifikacji, co stanowi utrudnienie dla większości metod stosowanych w biologii molekularnej.

Celem ogólnym niniejszej pracy była optymalizacja technologii CRISPR-Cas9 w kontekście celowania w regiony sekwencji powtarzającej się. W swoich badaniach używałam systemu CRISPR-Cas9 oraz jego modyfikacji do indukowania pęknięć nici DNA w regionie sekwencji CAG w genie *HTT*. Te badania przyczyniły się do opracowania nowej strategii terapeutycznej, która skutkowała „bezszwowym" wycięciem sekwencji CAG. W celu dokładnego poznania efektów edycji w obrębie sekwencji powtarzającej się stworzyliśmy nową metodę o nazwie qEva-CRISPR. Ta metoda jest dedykowana do wykrywania specyficznych i niespecyficznych cięć oraz rozróżniania zdarzeń wynikających z różnych mechanizmów naprawy DNA. Ponadto praca doktorska obejmuje zagadnienia związane z tworzeniem nowych modeli komórkowych HD za pomocą technologii CRISPR-Cas9. W niniejszej pracy uzyskane linie komórkowe posłużyły do testowania cząsteczek terapeutycznych oraz zbadania niektórych aspektów patogenezy HD.

Wyniki uzyskane w ramach tej rozprawy doktorskiej przyczyniły się do usprawnienia technologii CRISPR-Cas9 w kontekście jej aktywności w regionach sekwencji powtarzającej się oraz do zwiększenia efektywności i specyficzności modyfikacji DNA w tym regionie. Ponadto badanie efektów edycji w obrębie sekwencji CAG zrodziło pytania o to jakie mechanizmy są zaangażowane w jej naprawę oraz jak nimi sterować? Stworzona w ramach doktoratu metoda qEva-CRISPR może służyć do określania specyficzności cięcia zaprojektowanych komponentów systemu CRISPR-Cas9, włączając w to testowanie nowych białek należących do rodziny CRISPR. Ponadto nowe modele komórkowe umożliwią wykrycie bezpośredniego efektu ekspansji powtórzeń na fenotyp i funkcje komórki oraz pozwolą zbadać różne aspekty patogenezy HD.

SUMMARY

Polyglutamine disorders (polyQ) are rare neurodegenerative genetic diseases caused by the expansion of CAG repeats in associated genes. Huntington's disease (HD) occurs most frequently among polyQ disorders. Although the *HTT* gene responsible for this disease has been well characterized, aspects of HD pathogenesis are still unknown. Therefore, it is one of the reasons for the lack of potential therapy.

The CRISPR-Cas9 technology gives a broad range of opportunities to create new cellular models and establish potential therapies. Nevertheless, applying the CRISPR-Cas9 technology to explore repetitive sequences is still challenging due to a long CAG sequence which can form secondary structures of DNA and RNA, and *slipped strands*. Therefore, long repetitive stretches are problematic for the majority of methods used in molecular biology.

The doctoral thesis aimed to optimize the CRISPR-Cas9 technology in the context of targeting repetitive sequences. In my research, I used the CRISPR-Cas9 system and its modifications to induce DNA breaks in sequences flanking CAG repeats and within a repetitive sequence in the *HTT* gene. These studies have contributed to establishing a new potential therapeutic approach that caused "seamless" CAG repeat excision. To better explore the editing outcome within the CAG repeat sequence, we created a new method in the genome engineering field called Eva-CRISPR. This method is dedicated to detecting specific and unspecific cuts and distinguishing the DNA repair events. Moreover, the doctoral thesis included issues related to CRISPR-Cas9-based creating new cellular models of HD. In this work, we utilized new cell lines to test the therapeutic potential of drugs and to explore some aspects of HD pathogenesis.

The results obtained from these studies contributed to improving the CRISPR-Cas9 technology in the context of its activity within repetitive sequences and to increasing the efficacy and specificity of DNA modification in the *HTT* gene. Moreover, exploring editing effects within CAG repeats tract arose new questions: which DNA repair mechanisms are involved in repetitive sequences repair and how to manipulate them? A new method established during Ph.D. studies can be utilized to determine the cleavage specificity of CRISPR-Cas9 components, including new proteins belonging to the CRISPR family. New cellular HD models can be utilized to explore the effect of the CAG repeats expansion on cellular function and phenotype. Moreover, they are necessary to study various aspects of HD pathogenesis.

WPROWADZENIE

Choroby poliglutaminowe

Choroby poliglutaminowe (poliQ) są rzadkimi chorobami genetycznymi wywoływanymi ekspansją sekwencji powtórzeń CAG, które kodują glutaminę. Znanych jest dziewięć chorób poliQ, które są spowodowane nadmiernym wydłużeniem sekwencji mikrosatelitarnej CAG w regionach kodujących pojedynczych genów. Do tych zaburzeń zaliczamy: chorobę Huntingtona (ang. Huntington's disease, HD), ataksje rdzeniowo-móżdżkowe typu 1, 2, 3, 6, 7, 17 (ang. spinocerebellar ataxias, SCAs), zanik jądra zębatego, jądra czerwiennego, gałki bladej i jądra niskowzgórzowego (ang. dentatorubral-pallidoluysian atrophy, DRPLA) i rdzeniowo-opuszkowy zanik mięśni (ang. spinal and bulbar muscular atrophy, SBMA). Każda z dziewięciu chorób poliQ ma swój własny gen sprawczy oraz inny próg liczby powtórzeń, które wywołują chorobę (Tabela 1.) [1][2][3]. Te zaburzenia są dziedziczone w sposób autosomalny dominujący z wyjątkiem SBMA, które jest wywołane mutacją w genie kodującym receptor androgenowy znajdujący się na chromosomie X. Częstość występowania chorób poliQ szacuje się na około 1–10 przypadków na 100 000 ludzi. Spośród tych zaburzeń, HD i SCA3 mają najwyższą częstość występowania na świecie, inne choroby dominują w określonych obszarach geograficznych. DRPLA występuje głównie w Japonii, zaś SBMA występuje z dużą częstością w Finlandii [4][5].

Skrót nazwy choroby	**Gen**	**Lokalizacja**	**Długość sekwencji CAG**		
			Normalna	**Pośrednia**	**Patologiczna**
DRPLA	*ATN1*	12p13.31	7-35	35-47	49-88
HD	*HTT*	4p16.3	10-26	27-35	36-250
SBMA	*AR*	Xq12	5-34	35-46	37-70
SCA1	*ATXN1*	6p22.3	6-35	36-38	39-91
SCA2	*ATXN2*	12q24.12	14-31	27-33	33-500
SCA3	*ATXN3*	14q32.12	11-44	45-59	60-87
SCA6	*CACNA1A*	19p13.13	4-18	19	20-33
SCA7	*ATXN7*	3p14.1	4-19	28-35	34-460
SCA17	*TBP*	6q27	25-41	41	46-55

Tabela 1. Charakterystyka chorób poliQ.

Ekspansja powtórzeń pojawia się w dzielących i niedzielących się komórkach i jest zależna od rodzaju tkanki i choroby [6][7]. Nadmiernie wydłużona sekwencja CAG (powyżej 35 powtórzeń) tworzy II-rzędowe struktury, takie jak struktury typu spinki do włosów (ang. hairpin) lub tzw. *slipped strands* [6]. Te struktury DNA są zaangażowane w regulację funkcji komórkowych, takich jak organizacja chromatyny, ekspresja genów, jak również naprawa DNA [8][6][9]. Podczas

syntezy białka wydłużone sekwencje powtórzeń CAG są tłumaczone na szereg reszt glutaminowych, które tworzą ciąg poliQ. Nagromadzenie białek poliQ może osłabiać i uszkadzać mitochondria, białka opiekuńcze, oraz układ ubikwityna-proteasom [10][11][12]. W konsekwencji, te zagregowane białka poliQ znajdują się w zdegenerowanych neuronach w móżdżku, pniu mózgu i rdzeniu kręgowym [13]. Dlatego też różne białka, które zawierają ciąg poliQ ostatecznie prowadzą do dysfunkcji i degeneracji określonych subpopulacji neuronów[14]. Objawy chorób poliQ związane są głównie z upośledzeniem funkcjonowania układu nerwowego. W większości przypadków nasilają się one u chorych w średnim wieku i ich pogorszenie postępuje aż do śmierci pacjenta. W trakcie przekazywania powtórzeń kolejnym pokoleniom ich liczba prawie zawsze wzrasta, przez co objawy chorobowe mogą ujawnić się wcześniej i są silniejsze (antycypacja genetyczna) [7].

Choroby poliQ są obecnie nieuleczalne, można jedynie w ograniczony sposób niwelować ich objawy. W eksperymentalnej terapii tych chorób wykorzystywano różne strategie terapeutyczne, takie jak degradacja transkryptu za pomocą technologii antysensowych oligonukleotydów (ang. antisense oligonucleotide, ASO), jak również interferencji RNA (ang. RNA interference, RNAi) [15][16][17][18]. Oprócz tego, do manipulacji na poziome DNA w kontekście tych chorób były używane pierwsze systemy do edycji genomów, takie jak nukleazy z motywem palca cynkowego (ang. Zinc-finger nucleases, ZFNs) oraz TALENs (ang. transcription activator- like effector nucleases)[19][20]. Obecnie coraz częściej sięga się po nowe narzędzie do edycji genomu, jakim jest technologia CRISPR-Cas9 (ang. clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR associated protein 9 (Cas9)) [21][22].

Technologia CRISPR-Cas9

CRISPR-Cas jest systemem odpornościowym prokariontów i chroni je przed patogenami poprzez cięcie egzogennego DNA za pomocą nukleaz. Właściwości tego naturalnego systemu do niszczenia specyficznej sekwencji zostały wykorzystane do edycji genomów innych gatunków w technologii CRISPR-Cas9 [23][24]. CRISPR-Cas9 składa się z dwóch komponentów: białka Cas9 i krótkiego RNA, tzw. gRNA (ang. guide RNA). Około 20 nt sekwencja gRNA łączy się komplementarnie z celem po uprzednim rozpoznaniu przez nukleazę Cas9 sekwencji PAM (ang. Protospacer Adjacent Motif). Następnie białko Cas9 dokonuje cięcia około trzech nukleotydów powyżej motywu PAM. Efektem cięcia są podwójne pęknięcia nici DNA (ang. Double Strand Breaks, DSBs), które są naprawiane przez podatną na błędy ścieżkę scalania niehomologicznych

końców DNA (and. Non-Homologous End Joining, NHEJ) lub przy jednoczesnym dostarczeniu matrycy donorowej, mniej wydajną ale bardziej specyficzną naprawę homologiczną (ang. Homology Directed Repair, HDR). Efektem działania mechanizmu NHEJ są bezszwowe połączenie pęknięć nici DNA lub mutacje typu insercja/delecja (ang. insertion/deletion mutations, indels), a mechanizmu HDR synteza fragmentu DNA na podstawie matrycy donorowej w miejsce DSBs [24][25]. gRNA może tolerować pojedyncze niedopasowania i łączyć się z innymi regionami w genomie na zasadzie częściowej komplementarności, co warunkuje powstawanie mutacji typu indel w innych od przewidywanych sekwencjach (off-target). Białko Cas9 z bakterii S*treptococcus pyogenes* (*Sp*Cas9) jest najczęściej używanym białkiem w technologii CRISPR-Cas9. Motywami PAM dla tej nukleazy są kanoniczny PAM NGG i niekanoniczy PAM NAG (N oznacza dowolny nukleotyd). Jedną z wad tego białka jest to, że końcowym efektem jego działania mogą być tępe końce w miejscu DSBs, które dają mniejsze możliwości precyzyjnej naprawy. Oprócz tego *Sp*Cas9 jest dosyć duże (~1368 aminokwasów), nie jest wystarczająco specyficzne- daje dużo efektów ubocznych, takich jak cięcia poza celem, jak również może powodować rearanżacje chromosomów [25][26][27]. Wyżej wymienione cechy skłaniają do poszukiwania nowych białek oraz modyfikacji białka *Sp*Cas9, które pozwoliłyby usunąć te ograniczenia. Jednym z takich białek jest nukleaza Cas12a (poprzednio Cpf1), która jest mniejsza od *Sp*Cas9 (~1300 aminokwasów) i tnie sekwencję docelową około 17 i 23 nt od motywu PAM. Ponadto Cas12a oddziałuje z innymi motywami PAM, takimi jak TTV lub TTTV (V, to nukleotydy A, C albo G) oraz generuje lepkie końce w miejscu cięcia [28][29][30].

Jedną z modyfikacji dzikiej nukleazy *Sp*Cas9 jest białko nikaza Cas9 (ang. Cas9 nickase, Cas9n). Ta nukleaza ma mutację w jednej z domen tnących, przez co ma zdolność do cięcia tylko jednej nici DNA. Pojedyncze pęknięcia nici DNA (ang. Single Strand Breaks, SSBs) są naprawiane z większą precyzją (nawet 1500- krotnie mniej efektów niespecyficznych), a zastosowanie dwóch gRNA po przeciwległych stronach od planowanego miejsca modyfikacji może powodować pożądane delecje, przy jednoczesnym zmniejszeniu efektów off-target [31][32][33].

Kolejną modyfikacją dzikiej nukleazy *Sp*Cas9 jest białko *Sp*Cas9 z mutacją w dwóch domenach tnących (RuvC oraz HNH), tzw. dead Cas9 (dCas9). Kompleksy gRNA/dCas9 nie mają zdolności do cięcia sekwencji docelowych, a jedynie do ich wiązania [34][35].

Obecnie technologia CRISPR-Cas9 jest powszechnie używana, między innymi do: (I) tworzenia nowych modeli komórkowych i zwierzęcych z pożądanymi mutacjami/ zmianami w konkretnej lokalizacji w genomie [36][37], (II) śledzenia procesów biologicznych, które zachodzą w komórce [38][39], (III) wielkoskalowych badaniach funkcjonalnych (CRISPR screening) [40][41], jak również w (IV) badaniach przedklinicznych i potencjalnych terapiach dla wielu chorób genetycznych [42][43].

Technologia CRISPR-Cas9 w potencjalnej terapii chorób poliQ

Badania z użyciem technologii CRISPR-Cas9 w terapii chorób poliQ są wciąż na wczesnym etapie rozwoju przedklinicznego. Jak dotąd, system CRISPR-Cas9 został wykorzystany do obniżenia poziomu ekspresji *HTT* oraz do modyfikacji długości regionów mikrosatelitarnych w HD i SCA3 [44][22][45][46][47][21][48].

W pierwszym podejściu, wykorzystującym strategię CRISPRi (ang. CRISPR interference) obniżono o około 50% poziom huntingtyny poprzez użycie białka dCas9 i gRNA, które było komplementarne do miejsca startu transkrypcji genu *HTT* [44]. Inne strategie terapeutyczne polegały na inaktywacji genów poprzez nie allelo- selektywne lub allelo-selektywne usunięcie sekwencji powtarzającej się. W nie allelo-selektywnych strategiach używano Cas9 z parą gRNA, które były komplementarne do sekwencji 5' powyżej i 3' poniżej sekwencji CAG w genach *HTT* i *ATXN3*. Efektem tych podejść było usunięcie sekwencji CAG razem z fragmentami sekwencji oskrzydlającymi ciąg, co spowodowało przesunięcie ramki odczytu i powstanie przedwczesnego kodonu STOP na obu allelach. Nie allelo-selektywne wycięcie sekwencji powtarzającej się spowodowało około 80% wyciszenie huntingtyny w komórkach HEK 293T oraz całkowite wyciszenie *ATXN3* w indukowanych komórkach pluripotencjalnych (ang. induced pluripotent stem cells, iPSCs) i macierzystych komórkach neuronalnych (ang. neural stem cells, NSC) po edycji [46][21]. Podobne podejście z wykorzystaniem Cas9 odniosło również sukces w transgenicznym modelu mysim HD Q140, powodując redukcję poziomu huntingtyny, osłabienie patologii i poprawę funkcji motorycznych [49].

W podejściu allelo-selektywnym wykorzystano naturalnie występujące w sekwencji genu *HTT* polimorfizmy pojedynczych nukleotydów (ang. Single Nucleotide Polymorphisms, SNPs), które tworzyły motywy PAM w regionach 3' i 5' oskrzydlających powtórzenia na jednym bądź obu allelach. Kolejnym krokiem było zaprojektowanie dwóch gRNA oddziałujących z tymi motywami, które zastosowane w tym samym czasie generowały delecję ciągów CAG tylko na allelu z mutacją [22][46]. To podejście zostało przetestowane na fibroblastach, które pochodziły od pacjentów cierpiących na HD i powodowało obniżenie poziomu zmutowanej huntingtyny o około 50% [46]. Kolejne badania z użyciem tej strategii udowodniły nawet całkowite wyciszenie *HTT* w fibroblastach z sekwencją 44 powtórzeń CAG na zmutowanym allelu [22].

Powyższe strategie, które polegają na obniżeniu poziomu transkryptu *HTT* i wycięciu sekwencji CAG mają duży potencjał w leczeniu chorób poliQ. Dotychczasowe próby przedkliniczne potwierdzają wykonalność tych podejść, jednak mają one swoje ograniczenia. Potrzebne są dodatkowe badania i modyfikacje aby poprawić ich specyficzność.

Metody określania wydajności edycji

Znanych jest kilka metod do określenia efektów edycji wywołanych przez system CRISPR-Cas9 oraz oceny aktywności zaprojektowanych komponentów tego systemu. Te metody opierają się głównie na amplifikacji sekwencji, w której znajduje się przewidziane miejsce cięcia i rozdziale powstałych produktów w żelach agarozowych lub poliakrylamidowych. Ponadto można wyróżnić metody oparte na wykorzystaniu sond, jak również różnego typu platform do sekwencjonowania następnej generacji (NGS). Najczęściej używanymi metodami opartymi na PCR są metody enzymatyczne, takie jak T7E1 (ang. T7 endonuclease 1) i Surveyor assay [50][51]. Te metody wykorzystują cięcie heterodupleksów, które powstają na skutek łączenia się zmodyfikowanej i niezmodyfikowanej nici DNA. Niedopasowania pomiędzy tymi nićmi są cięte przez enzym, co powoduje pojawienie się specyficznego wzoru prążków podczas rozdziału w żelu. Efektywność edycji jest mierzona na podstawie porównania intensywności sygnału z prążków, które powstały po edycji w porównaniu do niezmodyfikowanej sekwencji.

Kolejnymi metodami, które opierają się na amplifikacji sekwencji, w której znajduje się przewidziane miejsce cięcia dla systemu CRISPR-Cas9 są IDAA (ang. Indel Detection by Amplicon Analysis) i TIDE (ang. Tracking of Indels by DEcomposition) [52][53][54]. IDAA wykorzystuje rozdział produktów PCR w elektroforezie kapilarnej. W tym przypadku nie jest znana sekwencja analizowanego fragmentu, a jedynie jego długość. Na tej podstawie można określić czy doszło do delecji/ insercji podczas naprawy DNA. Dla porównania, metoda TIDE umożliwia poznanie sekwencji DNA po edycji (w tej metodzie porównywane są elektroforegramy z sekwencjonowania Sangera przed i po edycji). Na tej podstawie algorytm porównuje obie sekwencje i określa efektywność edycji dla puli zdarzeń po cięciu.

Inną kategorią są metody oparte na użyciu sond, których przykładem jest ddPCR (ang. droplet digital PCR) [55]. W jego klasycznym wariancie występują dwie sondy, z czego jedna z nich jest komplementarna do miejsca cięcia, a kolejna jest sondą kontrolną, która jest oddalona od miejsca komplementarnego do sekwencji gRNA. Mutacje typu indel uniemożliwiają wiązanie się sondy do sekwencji, co skutkuje brakiem sygnału z sondy komplementarnej do miejsca cięcia. Ta metoda jest wykorzystywana do sprawdzenia efektywności edycji, jednak nie obrazuje ona sekwencji mutacji w miejscu cięcia. Ten problem może być rozwiązany przez sekwencjonowanie następnej generacji (ang. next-generation sequencing, NGS). Długość odczytów w tej metodzie, zależy od rodzaju platformy. Generalnie odczyty uzyskane z NGS są dość krótkie (~ 300–700 bp), co czasami przeszkadza w wykryciu dużych delecji. Odpowiedzią na to może być jednocząsteczkowe sekwencjonowanie DNA w czasie rzeczywistym (ang. single molecule real-time, SMRT), które zapewnia odczyty o średniej długości około 8,5 kB [56]. Poza tym, znane są

platformy typu Pac-BIO i Nanopore, które generują odczyty powyżej 10 kB. Jednak wraz ze wzrostem długości odczytu może spadać jego jakość [57][58].
Funkcjonalna analiza efektywności cięcia dla zaprojektowanych gRNA przewiduje również analizę niespecyficznych cięć. Obecnie znane są programy bioinformatyczne do przewidywana potencjalnych miejsc cięcia na podstawie komplementarności cząsteczki gRNA do innych miejsc w genomie. Przykładem takich programów są Cas-OFFinder i CRISPOR [59][60]. Na ich podstawie bez wykonywania eksperymentów można wyszukać potencjalne sekwencje typu off-target z 0 do 4 niedopasowaniami oraz określić ich lokalizację w genomie.

Wszystkie wymienione powyżej metody mogą być również zastosowane do analizy niepożądanych cięć, pod warunkiem, że jest znana ich lokalizacja w genomie. Na polu inżynierii genetycznej wykorzystywane są też metody do wykrywania efektów edycji, które mogą pojawić się w innych miejscach od tych proponowanych przez narzędzia bioinformatyczne. Te metody to różne warianty NGS np. sekwencjonowanie całego eksomu (ang. Whole Exome Sequencing, WES), sekwencjonowanie całego genomu (ang. Whole Genome Sequencing, WGS), BLESS (ang. breaks labeling, enrichment on streptavidin and next-generation sequencing) i Digenome-seq [61][62][63][64][51].

Nowe modele komórkowe chorób poliQ

Pomimo tego, że są znane mutacje odpowiedzialne za pojawienie się chorób poliQ, nadal nie są znane wszystkie aspekty patogenezy tych chorób, takie jak niestabilność somatyczna sekwencji CAG oraz rola białek poliQ w mózgu dorosłego człowieka [65].
Technologie CRISPR-Cas9 i CRISPR-Cas9n stwarzają możliwość generowania linii komórkowych z pożądaną zmianą w konkretnej lokalizacji w genomie na podstawie matrycy donorowej, która jest dostarczana razem z komponentami tych systemów do komórki. Umożliwia to, między innymi, badanie zależności wprowadzonej zmiany na fenotyp i procesy komórkowe [66][67].
Matryce donorowe mogą występować w postaci jednoniciowych oligonukleotydów (ang. single-stranded oligodeoxynucleotides, ssODNs) lub dwuniciowego DNA (plazmidy). Oprócz pożądanej sekwencji zawierają one również ramiona homologiczne, które są komplementarne do sekwencji otaczającej miejsce modyfikacji. W kontekście chorób poliQ naprawa DSBs na podstawie matrycy donorowej umożliwia tworzenie modeli komórkowych z wybraną długością sekwencji CAG [68].
W literaturze możemy odnaleźć kilka modeli komórkowych chorób poliQ, takich jak: HEK 293T i iPSCs, które zostały zmodyfikowane systemami CRISPR-Cas9 i CRISPR-Cas9n

[66][68][67][69][70][71][21]. Przykładem może być linia HEK 293T z wydłużoną sekwencją ciągu powtórzeń CAG (150 powtórzeń) w genie *HTT* oraz linia z delecją powtórzeń, która spowodowała inaktywację huntingtyny. Zostały one stworzone przy użyciu dzikiego białka Cas9 i gRNA celującego w sekwencję 5' powyżej sekwencji powtarzającej się oraz matrycy donorowej w postaci plazmidu [66]. Komórki HEK 293T nie są typowym modelem kojarzonym z chorobami poliQ, niemniej jednak są one proste w hodowli i w łatwy sposób można do nich wprowadzić cząsteczki terapeutyczne, co daje im przewagę nad innymi typami komórek.

Kolejnym modelem są iPSCs. Znanych jest kilka przykładów, w których technologie CRISPR-Cas9 i CRISPR-Cas9n zostały wykorzystane do stworzenia izogenicznych modeli iPSCs z prawidłową długością sekwencji ciągu powtórzeń CAG oraz modelu z wydłużoną sekwencją CAG. Jednym z nich jest linia izogeniczna wyprowadzona z komórek, które zawierały 18/180 powtórzeń CAG, w której użyto system CRISPR-Cas9n do wykonania dwóch cięć 5' powyżej sekwencji CAG. Matrycą donorową w tym podejściu był plazmid z ramionami homologicznymi i kasetą, która zawierała gen reporterowy GFP oraz gen oporności na puromycynę. Obecność tej kasety ułatwiła selekcję komórek z pożądaną zmianą [67]. Nowa linia została zróżnicowana do prekursorów neuronalnych i neuronów, w których zaobserwowano odwrócenie fenotypu HD. Wykorzystanie matrycy donorowej z genem oporności na neomycynę zostało użyte do stworzenia kolejnego modelu komórkowego z sekwencją 97 powtórzeń, która zawierała interrupcje CAA w genie *HTT*. W tych komórkach system CRISPR-Cas9 wygenerował DSBs w regionie 5' powyżej powtórzeń. Na podstawie matrycy donorowej zostały zsyntetyzowane nowe sekwencje 97 powtórzeń CAG w miejscach pierwotnych sekwencji na obu allelach (17/72 powtórzeń) [68].

W kontekście innych chorób poliQ technologia CRISPR-Cas9 została wykorzystana do stworzenia linii izogenicznych iPSCs z prawidłową długością sekwencji ciągu powtórzeń CAG w genie *ATXN2* (ta sama strategia jak w przypadku *HTT*) oraz z delecją tej sekwencji w genie *ATXN3* [69]–[71][21]. W przypadku linii z nieaktywną ataksyną 3 system CRISPR-Cas9 został wykorzystany do wygenerowania dwóch cięć w sekwencjach oskrzydlających powtórzenia CAG, co skutkowało ich delecją wraz z fragmentami sekwencji oskrzydlającej, a tym samym inaktywacją *ATXN3*.

CEL PRACY

HD to nieuleczalna choroba genetyczna, która występuje najczęściej ze wszystkich zaburzeń poliQ. Mimo że gen odpowiedzialny za rozwój choroby znany jest już od prawie 30 lat wciąż wiele aspektów patogenezy HD pozostaje niewyjaśnionych. Jak dotąd, brak też skutecznej metody terapii HD, a w ostatnim okresie kilka obiecujących badań klinicznych zakończyło się niepowodzeniem.

Technologia CRISPR-Cas9 stwarza ogromne możliwości, zarówno na polu tworzenia lepszych modeli badawczych, jak również opracowywania nowych podejść terapeutycznych dla chorób poliQ. Izogeniczne modele komórkowe, zawierające różną długość ciągu powtórzeń CAG przy identycznym tle genetycznym są niezbędne do wiarygodnych badań patogenezy chorób poliQ na poziomie DNA, RNA i białka. Wykorzystanie systemu CRISPR-Cas9 do badania sekwencji powtarzających się stanowi duże wyzwanie technologiczne. Wydłużona sekwencja CAG jest problematyczna dla większości metod stosowanych w biologii molekularnej, m.in. poprzez zdolność do tworzenia struktur drugorzędowych czy indukowania poślizgu polimerazy podczas amplifikacji. Ponadto, generowanie pęknięć DNA przez nukleazy Cas9 w obrębie sekwencji powtarzających się może angażować niekanoniczne ścieżki naprawy DNA i w rezultacie prowadzić do powstawania nietypowych produktów.

Celem ogólnym niniejszej pracy była optymalizacja technologii CRISPR-Cas9 w kontekście celowania w regiony sekwencji powtarzającej się. Cel ten realizowałam poprzez zastosowanie technologii edycji genomu do badania choroby Huntingtona, która jest przykładem chorób wywoływanych ekspansją powtórzeń CAG. System CRISPR-Cas9 został wykorzystany do modyfikacji regionu genu *HTT*, zawierającego wydłużony ciąg powtórzeń CAG.

Powyższy cel został osiągnięty przez realizację następujących zadań szczegółowych:

1. Opracowanie nowego podejścia terapeutycznego dla HD z wykorzystaniem nikazy Cas9
2. Stworzenie metody do oceny efektywności edycji, również w obrębie sekwencji mikrosatelitarnych
3. Stworzenie izogenicznych modeli komórkowych HD, służących do badania patogenezy i skuteczności potencjalnych terapeutyków

OMÓWIENIE PUBLIKACJI

Precise Excision of the CAG Tract from the Huntingtin Gene by Cas9 Nickases

W podejściach terapeutycznych, które zostały opisane w literaturze dotyczącej chorób poliQ używano dzikiego białka *Sp*Cas9, którego efektem działania były DSBs w regionie sekwencji CAG [22][45][46][49][21]. To białko często powoduje niespecyficzne cięcia poza pożądanym obszarem oraz rearanżacje chromosomowe, szczególnie jeżeli miejsca cięcia znajdują się blisko sekwencji mikrohomologicznej [72][33]. Dlatego w swoich badaniach opracowałam strategię, która polegała na indukcji SSBs przez system CRISPR-Cas9n w obszarze okalającym sekwencję powtórzeń CAG. Podczas replikacji dwa SSBs mogą być naprawiane przez mechanizmy naprawy DNA odpowiedzialne za naprawę DSBs, co skutkuje wydajniejszą edycją. Poza tym, istnieje małe prawdopodobieństwo, że dwa SSBs będą znajdowały się w bliskiej odległości od siebie w innych regionach genomu, co zmniejsza ryzyko pojawienia się niespecyficznych cięć.

W regionach bezpośrednio oskrzydlających powtórzenia CAG w modelu HD występuje tylko jedna kanoniczna sekwencja PAM, przez co zaprojektowanie specyficznych gRNA, dla których przewidywane miejsca cięcia występowałyby w regionie powtórzeń jest trudne. Trójnukleotyd CAG stanowi niekanoniczny PAM dla SpCas9 (NAG), jednak celowanie bezpośrednio w region powtórzeń może dawać duży efekt niespecyficznego cięcia. Zaprojektowałam cztery gRNA celujące w regiony 5’ powyżej (gRNA1, gRNA3) i 3’ poniżej (gRNA4) powtórzeń CAG oraz w obrębie tej sekwencji (gRNA2).

Rycina 1. gRNA celujące w sekwencję genu *HTT*. Na zielono zaznaczone są sekwencje PAM.

Efektywność edycji dla tych gRNA wynosiła od 12% do 40%. Następnie połączyłam gRNA w pary (gRNA1+gRNA4, gRNA3+gRNA4, gRNA2+gRNA4), które w tym samym czasie wprowadziłam do komórek HEK 293T (16/17 CAG). Pary gRNA, które składały się z gRNA1 i gRNA4, gRNA3 i gRNA4, powodowały precyzyjne, bezszwowe wycięcie sekwencji ciągu CAG (delecja 107 nt i 125 nt) w komórkach. Parę gRNA (gRNA1 i gRNA4), która charakteryzowała się najlepszym działaniem wybrałam do analiz w fibroblastach HD z różną długością sekwencji powtórzeń w *locus HTT*: 21/44 CAG, 17/68 CAG i 21/151 CAG. Okazało się, że długość powtórzeń nie ma wpływu na efektywność ich wycinania. Analiza metodą RT-qPCR wykazała, że poziom skróconego

transkryptu HTT nie zmienił się w stosunku do komórek nietraktowaną nikazą Cas9. Ponadto metoda western blot pokazała około 70% wyciszenie białka huntingtyny po zastosowaniu powyższej strategii we wszystkich modelach komórkowych HD. Dla pary gRNA, którą zastosowaliśmy w poprzednim podejściu z nikazą Cas9 został sprawdzony efekt niespecyficznego cięcia. Przy pomocy programu CRISPOR wybrałam przewidywane miejsca w genomie, które na zasadzie podobieństwa sekwencji mogły być celem dla gRNA1 i gRNA4. Analiza metodą T7E1 nie wykazała edycji potencjalnych sekwencji docelowych w genach *TEX13A, ZFHX3*, *TJP2*, *FBXW7*.

Podsumowując, stworzyliśmy nowe podejście terapeutyczne oparte na indukcji SSBs, które powoduje wydajną i specyficzną inaktywację genu *HTT.*

Gene Therapy for Huntington's Disease Using Targeted Endonucleases

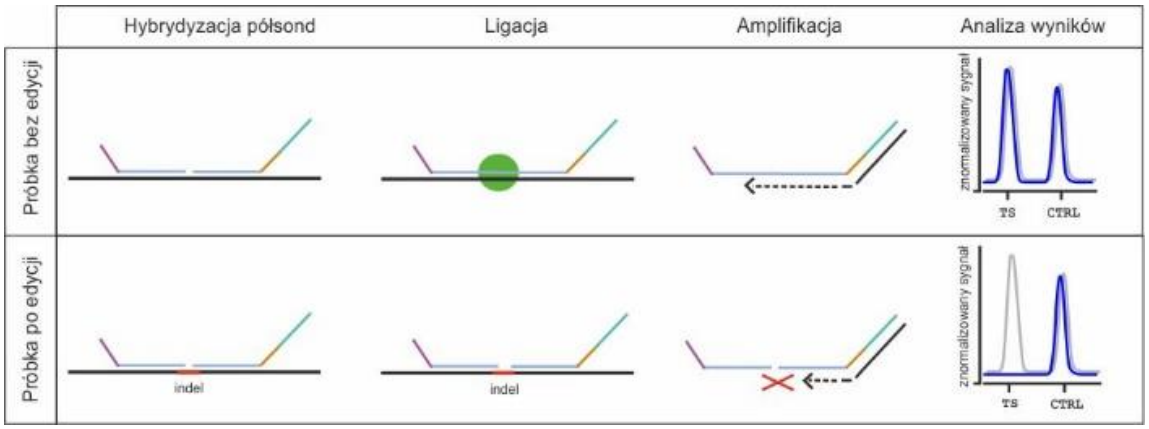
Edycja sekwencji CAG w genie *HTT* jest bardzo trudna. Problemy stwarza rozmieszczenie i niewielka liczba kanonicznych motywów PAM w obrębie sekwencji oskrzydlających ciąg. Poza tym, zwielokrotniona sekwencja CAG jest obecna w innych genach poliQ, dlatego powodowanie DSBs w regionie tej sekwencji może prowadzić do wielu niespecyficznych cięć i rearanżacji chromosomów. W celu zminimalizowania tych niepożądanych efektów opracowałam protokół, w którym proponuję edycję sekwencji powtórzeń poprzez indukcję SSBs, które są naprawiane z większą precyzją. Wynikiem takiej edycji genu *HTT* jest bezszwowa delecja sekwencji powtarzającej się razem z sekwencjami okalającymi ciąg, przesunięcie ramki odczytu i powstanie wczesnego kodonu STOP, co powoduje zahamowanie powstawania prawidłowego i zmutowanego białka huntingtyny. Ten protokół jest przeznaczony do edycji fibroblastów od pacjentów HD z różną długością sekwencji powtórzeń CAG. W tej strategii komponenty systemu CRISPR-Cas9n ulegają ekspresji z plazmidów. Każdy z nich posiada sekwencję jednego gRNA i nikazę Cas9. Ze względu na trudności w dostarczaniu cząsteczek terapeutycznych do fibroblastów HD proponujemy dostarczenie plazmidów z ekspresją nikazy Cas9 i gRNA3 na drodze elektroporacji, a następnie sortowanie komórek ze względu na obecność sygnału GFP (obecność sygnału GFP w komórkach świadczy o efektywności nukleofekcji). Jest to konieczna procedura, ponieważ pomimo wykorzystania elektroporacji efektywność transfekcji nie jest duża, pominięcie tego kroku może prowadzić do błędnej interpretacji efektywności edycji. Przy zachowaniu proponowanych procedur to podejście powinno skutkować około 70% obniżeniem poziomu huntingtyny.

Podsumowując, w tej publikacji przedstawiamy szczegółowy protokół do inaktywacji genu *HTT* przy pomocy systemu CRISPR-Cas9n.

qEva-CRISPR: a method for quantitative evaluation of CRISPR/Cas- mediated genome editing in target and off-target sites

Efektem edycji są mutacje typu indel, duże delecje lub insercje oraz bezszwowa ligacja nici DNA w miejscu cięcia. Taka heterogenna mieszanina wariantów utrudnia wiarygodną ocenę wydajności edycji, zwłaszcza w przypadku ilościowej charakterystyki zmian. Istnieje wiele metod do oceny efektywności działania systemów do edycji genomu. Te metody zazwyczaj wykorzystują amplifikację DNA i rozdział powstałych produktów w żelach agarozowych lub podczas elektroforezy kapilarnej [50][51][52]. Niektóre z nich nie wykrywają mutacji homozygotycznych, substytucji czy dużych delecji [52][53]. Poza tym, w przypadku metod, które wykorzystują PCR amplifikacja długich ciągów powtórzeń jest problematyczna (ślizganie polimerazy na powtórzeniach, tworzenie struktur II-rzędowych). Aby rozwiązać powyższe problemy, we współpracy z Zakładem Genetyki Molekularnej stworzyliśmy nową metodę o nazwie qEva-CRISPR, służącą do badania wydajności edycji w obrębie sekwencji powtarzającej się, jak również w obrębie innych sekwencji w genomie.

qEva-CRISPR to połilościowa metoda, która służy do wykrywania zmian wprowadzonych przez system CRISPR-Cas9 w miejscach docelowych, jak również w przewidywanych metodami bioinformatycznymi innych miejscach w genomie (off-target). Jest ona oparta na dobrze znanej i zwalidowanej metodzie MLPA (ang. Multiplex Ligation-dependent Probe Amplification), która wykorzystuje amplifikację sond komplementarnych do sekwencji docelowej [73]. W metodzie qEva-CRISPR każda z sond, składa się z dwóch półsond, dla których miejscem ligacji jest przewidziane miejsce cięcia dla systemu CRISPR-Cas9. Jeżeli w miejscu cięcia występuje mutacja typu indel, półsondy nie ulegną ligacji lub nie połączą się komplementarnie z sekwencją docelową. Skutkuje to brakiem matrycy w postaci sondy, a co za tym idzie brakiem jej amplifikacji. Sygnał z próbki po edycji jest normalizowany do średniego sygnału z kilku sond kontrolnych (sondy kontrolne są komplementarne do miejsc w genomie, które nie uległy edycji), a następnie odnoszony do znormalizowanego sygnału pochodzącego od tej sondy w próbce bez mutacji.



Rycina 2. Uproszczony schemat działania metody qEva-CRISPR. TS oznacza sondę komplementarną do celu, zaś CTRL oznacza sondę kontrolną.

Aby sprawdzić, czy qEva-CRISPR ma zdolność do wykrywania efektów edycji stworzyliśmy układ sond, który wykrywał mutacje generowane przez system CRISPR-Cas9 w genie *TP53*. Oprócz tego zaprojektowaliśmy 8 sond kontrolnych i sondy nacelowane na potencjalne regiony off-target, co pozwoliło w jednej reakcji w tym samym czasie ustalić efektywność edycji oraz zbadać efekty niespecyficzne. Wydajność edycji w dużej mierze zależy od ilości komponentów systemu CRISPR-Cas9, które są dostarczane do komórek. Wykorzystując ten sam układ sond wykazaliśmy pozytywną korelację między wydajnością edycji genu *TP53* w komórkach HCT116 i HeLa a ilością plazmidowego DNA, z którego ekspresji ulegały gRNA i Cas9.

W kolejnym etapie sprawdziliśmy czułość metody qEva-CRISPR poprzez analizę zdolności do wykrywania pojedynczych delecji, insercji i substytucji. W tym eksperymencie zaprojektowaliśmy sondy, które zawierały jednonukleotydowe niesparowania z sekwencją docelową i sprawdziliśmy czy ich sygnał zmienił się w stosunki do próbki kontrolnej. We wszystkich przypadkach zaobserwowaliśmy brak sygnału w próbkach zawierających zmodyfikowane sondy.

Aby porównać wyniki uzyskane metodą qEva-CRISPR z innymi metodami z danych literaturowych wybraliśmy dobrze zwalidowane gRNA dla genów *VEGFA*, *CCR5* i *EMX1*. Efektywność działania tych gRNA została wcześniej sprawdzona metodami T7E1 i NGS [31][61][42]. Przy zachowaniu tych samych warunków eksperymentalnych (metody dostarczania systemu i typ komórek) wykazaliśmy, że wyniki otrzymane metodą qEva-CRISPR były porównywalne do wyników uzyskanych innymi metodami. Ponadto, qEva-CRISPR pozwoliła w sposób ilościowy w jednej reakcji wykryć efekty edycji 3 różnych genów oraz określić poziom niespecyficznych cięć.

Potencjał qEva-CRISPR do wykrywania zdarzeń wynikających z naprawy DNA został również przetestowany na „trudnych” sekwencjach ze zwielokrotnioną liczbą powtórzeń CAG. W tym

eksperymencie zaprojektowałam gRNA celujące w regiony 5’ powyżej i 3’ poniżej sekwencji powtarzającej się w genie *HTT.* Takie regiony są zazwyczaj bardzo trudne lub niemożliwe do analizy przez standardowe metody stosowane w biologii molekularnej m.in. ze względu na dużą zawartość par GC. W naszych eksperymentach zaprojektowaliśmy układ sond, który wykrywał efekty edycji m.in. bezpośrednio w zwielokrotnionej sekwencji CAG. Aby zwiększyć specyficzność tej strategii jedna z półsond była komplementarne do sekwencji okalającej ciąg, a druga półsonda była całkowicie komplementarna do sekwencji powtarzającej się. Eksperymenty przeprowadzone w czterech powtórzeniach biologicznych wyraźnie wykazały, że każdy z zastosowanych gRNA skutecznie indukował mutacje w regionie sekwencji CAG. Jednak efektywność edycji zależała od formy dostarczenia komponentów systemu CRISPR-Cas9 do komórek. Przy zastosowaniu dwóch metod dostarczania systemu CRISPR-Cas9 (wektory plazmidowe i kompleksy RNP) wykazaliśmy wyższą efektywność edycji po zastosowaniu wektorów plazmidowych, przy jednoczesnym zwiększonym udziale cięć niespecyficznych.

W erze edycji genomu bardzo ważną rolę odgrywa znajomość mechanizmów, które naprawiają SSBs i DSBs. NHEJ i HDR, to dwie główne ścieżki odpowiedzialne za naprawę DSBs w komórkach ludzkich. Efektem naprawy NHEJ w przeprowadzonych eksperymentach były mutacje typu indel, które powodowały obniżenie sygnału dla zaprojektowanych sond w genach *TP53*, *VEGFA*, *CCR5*, *EMX1* i *HTT*. Aby sprawdzić, czy qEva-CRISPR ma zdolność do wykrywania zdarzeń wynikających z naprawy HDR stworzyliśmy linie komórkowe z fragmentem sekwencji GFP w miejscu sekwencji CAG. Do stworzenia tych linii użyliśmy nikazę Cas9 oraz parę gRNA, które były komplementarne do sekwencji okalającej ciąg CAG. W tym eksperymencie zaprojektowaliśmy sondy, które wykrywały nowy fragment sekwencji GFP (HDR) takie, które wykrywały około 100 nt delecję (NHEJ), jak również niezmienioną sekwencję. Obecność pożądanego fragmentu sekwencji w komórkach HEK 293T przejawiała się wzrostem sygnału dla sondy komplementarnej do tej sekwencji. W zaprojektowanym układzie sond znaleźliśmy linie z fragmentem sekwencji GFP na obu allelach albo linie z sekwencją GFP na jednym allelu i delecją sekwencji CAG lub niezmienioną sekwencją na drugim allelu.

Podsumowując, stworzyliśmy nową metodę do jednoczesnego wykrywania specyficznych i niespecyficznych zmian w genomie, które są spowodowane przez system CRISPR-Cas9. qEva-CRISPR wykrywa każdy rodzaj mutacji, od pojedynczych substytucji aż po duże delecje i jej czułość nie zależy od typu mutacji. Eksperymenty, które przeprowadziliśmy z wykorzystaniem pięciu genów docelowych, włącznie z *HTT* oraz czterech linii komórkowych udowodniły, że qEva-CRISPR jest metodą specyficzną i czułą i może stanowić alternatywę dla obecnie stosowanych metod.

Generation of new isogenic models of Huntington's disease using CRISPR/Cas9 technology

W badaniach nad patogenezą chorób poliQ ważne jest posiadanie linii komórkowych z różną długością sekwencji ciągu powtórzeń CAG, które odpowiadają różnym wariantom choroby. Sekwencja, która posiada około 40 powtórzeń CAG jest kojarzona z najczęściej występującym wariantem patologicznym, zaś formy młodzieńcze HD posiadają więcej niż 60 powtórzeń CAG w genie *HTT* [74][75]. W przypadku iPSCs w literaturze możemy znaleźć linie izogeniczne wyprowadzone z komórek, które posiadały bardzo długie ciągi CAG oraz ciągi z interrupcjami CAA, które zazwyczaj nie występują u pacjentów z HD [67][68]. W celu lepszego poznania wpływu ciągów CAG na patogenezę HD stworzyliśmy nowe linie HEK 293T z patologicznie wydłużoną sekwencją ciągu CAG i izogeniczne linie iPSCs z prawidłową długością sekwencji powtarzającej się.

W swoich badaniach przetestowałam kilka strategii w celu wygenerowania linii z pożądaną długością sekwencji ciągu powtórzeń CAG w genie *HTT*. Pierwsze strategie opierały się na użyciu nikazy Cas9 i pary gRNA (gRNA1 i gRNA4) lub tylko jednego gRNA (gRNA4) oraz matrycy donorowej w postaci ssODN. Nie udało się dzięki nim otrzymać linii z wydłużoną sekwencją CAG. Kolejne nieudane strategie polegały na wykorzystaniu dzikiej nukleazy Cas9 oraz gRNA4 przy jednoczesnym dostarczeniu do komórek matryc donorowych w postaci ssODN lub plazmidu. W podejściu z dzikim białkiem Cas9 i gRNA3 oraz matrycą donorową w formie plazmidu udało się uzyskać linie z wydłużoną sekwencją CAG. Niestety każda z tych linii posiadała również dodatkowe mutacje w miejscu cięcia, które zmieniały ramkę odczytu. Dopiero wprowadzenie mutacji cichej w miejscu PAM w sekwencji matrycy donorowej spowodowało uzyskanie linii komórkowych z wydłużoną sekwencją CAG bez dodatkowych mutacji. Za pomocą tej strategii wyprowadziłam linie homozygotyczne HEK 293T z sekwencją 41, 53 i 84 powtórzeń CAG w genie *HTT*. W wyprowadzonych liniach komórkowych został sprawdzony poziom transkryptu HTT i białka oraz długość sekwencji ciągu CAG. Z przeprowadzonych badań wynika, że poziom transkryptu zmierzony metodą RT-qPCR koreluje z długością sekwencji powtórzeń CAG. Białka wykryte metodą western blot migrowały w żelach poliakrylamidowych zgodnie z długością domeny poliQ, a intensywność sygnału była odwrotnie skorelowana z długością sekwencji tego ciągu. Długość sekwencji powtarzającej się w genie *HTT* została potwierdzona za pomocą sekwencjonowania Sangera i odzwierciedlała długości sekwencji powtórzeń CAG w matrycach donorowych.

Te modele komórkowe posłużyły do sprawdzenia potencjału terapeutycznego dwóch cząsteczek siRNA oraz zbadania zjawiska, które polega na powstawaniu skróconego transkryptu HTT

zawierającego intron 1. siRNA celowały w sekwencję eksonu 1 w ramach nie allelo-selektywnej (siHTT) lub allelo –selektywnej (siRNA_A2) strategii, gdzie wykazały odpowiednio około 50% i 30% redukcję poziomu huntingtyny we wszystkich modelach. Powstawanie transkryptów wczesnego intronu 1 jest wiązane z zaburzeniami składania mRNA spowodowanymi nadmiernie wydłużoną sekwencją ciągu CAG w pierwszym eksonie *HTT*. Aby wykryć transkrypty wczesnego intronu 1 w komórkach HEK 293T z wydłużoną sekwencją powtarzającą się wykorzystałam metodę RT-PCR. Poziom transkryptów wczesnego intronu 1 był podwyższony w komórkach HEK 293T z sekwencją 41, 53 i 84 powtórzeń CAG w porównaniu do niezmodyfikowanych komórek HEK 293T (16/17 CAG).

Kolejnym modelem stworzonym w ramach tej pracy były izogeniczne linie iPSCs z prawidłową długością sekwencji ciągu powtórzeń oraz linia z nieaktywną huntingtyną. Wyjściowa linia, którą użyliśmy do wyprowadzenia nowych modeli posiadała sekwencje 19/~109 powtórzeń CAG w genie *HTT*. iPSCs charakteryzują się niską częstością występowania naprawy na drodze HDR, co skłoniło nas do przetestowania różnych strategii edycji. Pierwsza z nich polegała na użyciu wektorów ekspresyjnych z parą gRNA celującą w region 5’ powyżej (gRNA1) i 3’ poniżej (gRNA4) sekwencji powtarzającej się w połączeniu z nikazą Cas9. Matrycą donorową był ssODN bez lub z mutacją cichą w miejscu PAM. Te strategie nie były efektywne. W kolejnych podejściach używaliśmy białka Cas9 i gRNA3 w formie kompleksu RNP oraz matrycy donorowej w postaci ssODN. Ponadto, w tym eksperymencie zastosowaliśmy nokodazol, który zatrzymywał komórki w fazie S/M cyklu komórkowego. Zamiana plazmidu na kompleksy RNP oraz synchronizacja cyklu komórkowego spowodowały pojawienie się linii monoklonalnych z pożądaną zmianą. Jednak większość z nich oprócz zdarzeń wynikających z HDR posiadała również mutacje typu indel na drugim allelu. Podobna strategia, ale z zastosowaniem matrycy donorowej w postaci przeciętego plazmidu z mutacją cichą w miejscu PAM, okazała się efektywna i pozwoliła uzyskać pożądane linie. Spośród 131 kolonii uzyskano 8, które posiadały 19/19 powtórzeń CAG w genie *HTT*. Trzy spośród kolonii, które powstały na skutek HDR (C105, C39, C31.9) i jedna linia z nieaktywną huntingtyną (C37), która powstała w pierwszej strategii zostały scharakteryzowane pod kątem ekspresji prawidłowego transkryptu i białka. Analiza oparta na qPCR potwierdziła, że linie C105, C39 i C37 mają prawidłowy kariotyp, zaś linia C31.9 może posiadać amplifikację analizowanego regionu na chromosomie 4. Analiza RT-qPCR markerów *NANOG*, *OCT4* i *SOX2* wykazała prawidłowy, podobny poziom ich ekspresji we wszystkich wyprowadzonych liniach. Analizy mikroskopowe markerów pluripotencji *OCT4* i *NANOG* oraz białek powierzchniowych TRA 1–80 i TRA 1–60 również potwierdziły ich obecność w wyprowadzonych liniach.

Podsumowując, w tej pracy zaprezentowaliśmy różne strategie do generowania linii z wydłużoną i skróconą sekwencją CAG w genie *HTT*. Podejście z wykorzystaniem dzikiego białka Cas9 i gRNA3

z matrycą donorową w postaci zlinearyzowanego plazmidu okazało się najbardziej efektywne. Nowe linie HEK 293T i iPSCs zostały dobrze scharakteryzowane i posłużyły między innymi do zbadania potencjału terapeutycznego cząsteczek siRNA oraz zaburzeń składania mRNA.

PODSUMOWANIE I DYSKUSJA

Opracowanie nowego podejścia terapeutycznego dla HD z wykorzystaniem nikazy Cas9

W potencjalnej terapii chorób poliQ stosowane są różne podejścia terapeutyczne, których celem jest zmniejszenie ekspresji huntingtyny lub jej inaktywacja na poziomie DNA [44][22]. Funkcja huntingtyny nie jest do końca poznana, co spowodowało dwutorowy rozwój strategii terapeutycznych. Opierają się one na allelo- selektywnym lub nie allelo-selektywnym wyciszeniu/ inaktywacji tego białka. Badania z użyciem interferencji RNA oraz ASO na mysim modelu wykazały, że całkowita inaktywacja huntingtyny lub tylko jej zmutowanego allelu miały korzystny wpływ na fenotyp i funkcje motoryczne u myszy [76][77].

W związku z tym, w ramach swojej pracy skupiłam się na stworzeniu nowego podejścia terapeutycznego, które polegało na inaktywacji huntingtyny poprzez precyzyjne wycięcie sekwencji powtarzającej się w genie *HTT*. W tej strategii użyłam nikazy Cas9 w połączeniu z parą gRNA, oskrzydlających ciąg powtórzeń CAG. Białko nikaza Cas9 jest bardziej specyficzne w porównaniu z dziką nukleazą Cas9, przez co znacznie redukuje ryzyko niepożądanych cięć oraz rearanżacji chromosomowych. Efektywność działania systemu zależy od aktywności nukleolitycznej zaprojektowanych kompleksów gRNA/białko oraz odległości pomiędzy potencjalnymi miejscami cięcia dla zaprojektowanych gRNA. Wykazałam, że długość powtórzeń, a tym samym dystans pomiędzy potencjalnymi miejscami cięcia dla Cas9n-gRNA nie mają wpływu na efektywność ich wycinania. W przypadku allelu zawierającego 151 powtórzeń CAG region pomiędzy parą gRNA obejmował blisko 500 par zasad. To podejście spowodowało bezszwowe wycięcie sekwencji powtarzającej się razem z fragmentami sekwencji otaczającej ciąg. Nowa strategia wykazała około 70% obniżenie poziomu huntingtyny w komórkach pochodzących od pacjentów HD. Ponadto, niezaprzeczalną zaletą tej strategii było to, że nie wywołała ona efektów niespecyficznych w innych obszarach genomu.

Co ciekawe, po wycięciu sekwencji powtarzającej się zaobserwowaliśmy brak obniżenia poziomu transkryptu HTT. Najprawdopodobniej jest to spowodowane tym, że transkrypt może „uciekać" przed zjawiskiem degradacji transkryptów zawierających przedwczesny kodon STOP jakim jest

NMD [78][79]. Wycinanie powtórzeń z sekwencjami oskrzydlającymi zmieniło ramkę odczytu i spowodowało pojawienie się przedwczesnego kodonu STOP w eksonie 1 genu *HTT*. PTC pojawił się w pozycji zajmowanej przez EJC (ang. exon junction complex) (20-24 nt powyżej złączenia ekson/ ekson), przez co może być on maskowany podczas pierwszej rundy translacji. We wcześniejszych badaniach wykazano, że usunięcie systemem CRISPR-Cas9 sekwencji powtarzającej się spowodowało istotny spadek ekspresji transkryptu HTT w modelach komórkowych HD [22][80]. Oporność mRNA HTT na NMD, która jest obserwowana w moich badaniach może wynikać ze specyficznej lokalizacji PTC, który powstaje w eksonie 1 genu *HTT* na skutek wycięcia sekwencji powtórzeń. Zakładamy, że powstający transkrypt (bez ciągu powtórzeń CAG) nie powinien stanowić zagrożenia dla komórki, ponieważ wszystkie znane mechanizmy toksyczności RNA w HD są związane z obecnością zmutowanych powtórzeń [81][82].

Podsumowując, stworzyliśmy nowe podejście terapeutyczne oparte na inaktywacji huntingtyny przy użyciu systemu CRISPR-Cas9n. Niezaprzeczalną zaletą tego podejścia jest brak efektów niespecyficznego cięcia, co jest niezwykle pożądane przy tworzeniu potencjalnych terapii genetycznych.

Stworzenie metody do oceny efektywności edycji, również w obrębie sekwencji mikrosatelitarnych

Metodami, które umożliwiają wykrycie zmian spowodowanych przez technologię CRISPR-Cas9 w genomie są techniki oparte na enzymatycznym cięciu heterodupleksów tworzonych przez zmodyfikowany i niezmodyfikowany wariant genu- np. T7E1 i Surveyor assay [50][51]; sekwencjonowaniu Sangera- TIDE [53][54]; elektroforezie kapilarnej- IDAA [52]; ddPCR [55]; oraz na NGS [51]. Główną zaletą NGS jest zdolność do wykrywania wszystkich mutacji spowodowanych przez system CRISPR-Cas9, z drugiej jednak strony NGS jest nadal kosztowną i czasochłonną metodą (czas przygotowania biblioteki NGS oraz analiza danych), do której jest potrzebna specjalna aparatura. Metody oparte na ddPCR również wymagają specjalnego sprzętu, dzięki któremu próbka zostanie rozbita na tysiące kropli. Dla porównania pozostałe metody mają bardzo krótki czas analizy (około 1-2 dni) oraz nie wymagają specjalistycznego sprzętu. Jednak wykorzystują one amplifikację fragmentu sekwencji w którym znajduje się miejsce cięcia, co może wpłynąć na jakość wyników uzyskanych z „trudnych" sekwencji powtarzających się. Podczas prac nad edycją genu *HTT* pojawiło się wiele problemów związanych z analizą, m.in. heterogenność produktów PCR, która wynikała z poślizgu polimerazy i obecności

dwóch alleli z różną długością sekwencji ciągu CAG. W związku z tym, nawet próbka kontrolna (nietraktowana Cas9) była niejednorodna podczas rozdziału w żelach agarozowych, co utrudniło wiarygodną ocenę wydajności edycji genu *HTT* metodą T7E1. Ponadto poślizgi polimerazy w obrębie sekwencji CAG uniemożliwiły uzyskania odpowiedniej jakości elektroforegramów, a tym samym wpłynęły na błędne określenie wydajności edycji w metodzie TIDE.

Opracowana przez nas metoda qEva-CRISPR polega na amplifikacji pary sond przy pomocy uniwersalnych starterów, które są komplementarne do stałych fragmentów tych sond. Dzięki temu mogliśmy ocenić w sposób ilościowy wydajność edycji w *HTT* i efekt off-target dla kilku gRNA w jednej reakcji. Poza tym odpowiedni układ sond qEva-CRISPR pozwolił odróżnić zdarzenia wynikające z naprawy HDR od NHEJ w obrębie sekwencji CAG w *HTT*. Niewątpliwą zaletą qEva-CRISPR jest to, że wykrywa ona w jednej reakcji efekty edycji spowodowane przez system CRISPR-Cas9 w kilku miejscach w genomie. W odróżnieniu od metod, które polegają na amplifikacji regionu docelowego qEva-CRISPR wykrywa różne zdarzenia wynikające z naprawy DSBs, które mogą być zlokalizowane w różnych typach sekwencji. Pomimo licznych zalet nasza metoda ma też swoje wady. Jedną z nich jest to, że wykrywa ona tylko mutacje w sekwencjach na które zostały zaprojektowane sondy oraz może mieć problem z wykrywaniem mutacji, które występują z częstością poniżej 5%.

W moim nowym projekcie qEva-CRISPR jest używana do badania wariantów sekwencji, które powstały na skutek różnych mechanizmów naprawy DNA. Przy wyciszeniu wybranych genów naprawczych ta metoda pozwala określić udział konkretnego wariantu sekwencji w puli zdarzeń po edycji. W przyszłości qEva-CRISPR może być z powodzeniem używana do badania efektów edycji, które będą spowodowane przez nowe nukleazy.

Stworzenie izogenicznych modeli komórkowych HD, służących do badania patogenezy i skuteczności potencjalnych terapeutyków

Jednym z czynników, który wpływa na efektywność naprawy HDR jest odległość pomiędzy miejscem cięcia, a miejscem planowanej syntezy nowej sekwencji. W przypadku *HTT* niekanoniczna sekwencja PAM (CAG) tworzy sekwencję ciągu CAG, jednak jej użycie mogłoby doprowadzić do pojawienia się wielu pęknięć nici DNA w obrębie tej sekwencji, co doprowadziłoby do jej skrócenia. Jedynie kilka kanonicznych sekwencji PAM znajduje się w odległości od kilku do kilkudziesięciu nukleotydów od sekwencji ciągu CAG. Umożliwia to projektowanie gRNA komplementarnych do sekwencji okalających sekwencję CAG, co ogranicza

ryzyko wystąpienia efektów typu off-target. Ze względu na powyższe problemy tworzenie nowych modeli komórkowych HD było trudne i wymagało przetestowania wielu strategii celowania.

Stworzone w ramach pracy nowe modele komórkowe umożliwiają badania patogenezy HD oraz ułatwiają testowanie cząsteczek terapeutycznych. Były to linie HEK 293T z wydłużoną sekwencją CAG, izogeniczne linie iPSCs z prawidłową długością tej sekwencji oraz linia iPSC z nieaktywną huntingtyną. Do stworzenia powyższych modeli zostały użyte systemy CRISPR-Cas9 i CRISPR-Cas9n z matrycami donorowymi, które posiadały pożądaną długość sekwencji CAG. W ramach tej pracy udowodniliśmy, że strategia która opierała się na dzikim białku Cas9 oraz gRNA3 w połączeniu z matrycą donorową jaką był zlinearyzowany plazmid z mutacją cichą w sekwencji PAM okazała się najbardziej wydajna.

W literaturze możemy odnaleźć przykłady linii HEK 293T z wydłużoną sekwencją powtarzającą się (>80 powtórzeń CAG) [66]. W swoich badaniach jako pierwsi stworzyliśmy linie, które odpowiadają różnym wariantom patologicznym. Linie z sekwencją 41, 53 powtórzeń CAG, które występują najczęściej u pacjentów oraz formę młodzieńczą HD z sekwencją 84 powtórzeń CAG. Wszystkie linie są homozygotami pod względem sekwencji powtarzającej się. Ten układ jest pożądany ponieważ nie sprawia problemów w analizach długości sekwencji powtarzającej się metodami PCR, sekwencjonowaniem i metodą western blot.

Pomimo tego, że HEK 293T nie są komórkami neuronalnymi mogą być przydatne do badania wpływu długości sekwencji powtórzeń CAG na fenotyp komórki. Za pomocą mikroskopii elektronowej inne grupy badawcze wykazały, że inaktywacja huntingtyny w tych komórkach nie miała znaczącego wpływu na ich strukturę. Inny efekt powodowało wydłużenie sekwencji CAG. Ekspansja przyczyniła się do zmian morfologicznych mitochondriów, zwiększyła ich powiązanie z gładką i szorstką siateczką endoplazmatyczną oraz spowodowała nagromadzenie małych autolizosomów w cytoplazmie [66]. W naszych badaniach na komórkach HEK 293T z wydłużoną sekwencją powtarzającą się zostały wykazane zaburzenia składania mRNA, które polegały na powstaniu transkryptów wczesnego intronu 1 HTT. Ta nieprawidłowość wynika z obecności wydłużonej sekwencji CAG w pierwszym eksonie *HTT* i ostatecznie prowadzi do powstania krótkich, N-końcowych fragmentów tego genu. Oprócz tego te linie komórkowe okazały się świetnym modelem do testowania cząsteczek terapeutycznych, które wcześniej były stosowanie do degradacji transktyptów HTT na trudnych w hodowli komórkach od pacjentów HD [83].

Istnieje jednak wiele innych aspektów, które nie mogą być zbadane na tego typu modelach np. tło genetyczne HD i związane z tym zaburzenia transkryptomu. Stworzone przez nas izogeniczne linie iPSCs z prawidłową długością sekwencji CAG oraz linia z nieaktywną huntingtyną pozwalają

zbadać te nieprawidłowości. Poprzednio do tworzenia linii iPSCs były używane strategie, które opierały się na systemach CRISPR-Cas9 i CRISPR-Cas9n oraz matrycach donorowych jakimi były plazmidy z kasetami selekcyjnymi [67][68]. W tych podejściach mamy więcej technicznych aspektów, które mogą okazać się problematyczne w rutynowym tworzeniu modeli. Nasze linie iPSCs zostały dobrze scharakteryzowane, między innymi pod względem markerów pluripotencji oraz niespecyficznych cięć (WES). Te analizy nie wykazały żadnych nieprawidłowości. Oznacza to, że w przyszłości mogą być one zróżnicowane do np. prekursorów neuronalnych i neuronów, co ułatwi badania nad patogenezą HD.

Obecnie linie HEK 293T z wydłużoną sekwencją CAG są używane do sprawdzenia bezpośredniego efektu ekspansji powtórzeń na fenotyp i funkcje komórki w Zakładzie Inżynierii Genomowej i Biotechnologii Medycznej, jak również w innych zakładach w instytucie i jednostkach naukowych na świecie (np. the Francis Crick Institute w Londynie czy California Institute of Technology). Oprócz tego zostało rozwinięte zagadnienie związane z powstawaniem wczesnego intronu pierwszego HTT. Za pomocą opisanej strategii stworzyliśmy linie z interrupcjami CAA w sekwencji CAG, które destabilizują struktury II-rzędowe. Spowodowało to zahamowanie ekspresji nieprawidłowych transkryptów. Obecnie w Zakładzie Inżynierii Genomowej trwają badania nad poznaniem tego zjawiska.

Wyżej opisana strategia została wykorzystana do wyprowadzenia linii HEK 293T z sekwencją 35 powtórzeń CAG w genach *ATN1* i *ATXN3*. Te modele komórkowe jak również linie HEK 293T z sekwencją 41, 53 i 84 powtórzeń CAG w genie *HTT* są obecnie używane do badania mechanizmów naprawy DNA wewnątrz sekwencji powtarzających się.

Linie iPSCs zostały zróżnicowane do prekursorów neuronalnych HD i poddano je wnikliwej analizie transkryptomu. Ponadto, linia z nieaktywną huntingtyną będzie wykorzystana do zbadania funkcji tego białka w komórkach neuronalnych.

PERSPEKTYWY

Celowanie w sekwencje powtarzające się powoduje wycięcie, skrócenie lub wydłużenie tej sekwencji [22][84]. Dzięki poznaniu mechanizmów odpowiedzialnych za powstawanie tych zjawisk będzie można przewidzieć efekt edycji bądź nim sterować.

System CRISPR-Cas9 stwarza możliwość badania wpływu różnych czynników na naprawę DNA w konkretnej lokalizacji w genomie. Ta technologia pozwoli poznać mechanizmy odpowiedzialne za naprawę pęknięć nici DNA wewnątrz zmutowanej sekwencji CAG oraz przyczyni się do znalezienia sposobu kontrolowania procesów naprawczych. Moje obecne badania skupiają się na poznaniu ścieżek naprawy DNA uczestniczących w naprawie pęknięć nici DNA w obrębie sekwencji mikrosatelitarnych. W badaniach wykorzystuję analizy wielkoskalowe tzw. CRISPR interference screen oraz nukleazę Cas12a. Motywy PAM dla tej nukleazy znajdują się w sekwencjach okalających ciąg powtórzeń w genach poliQ takich jak: *ATN1*, *ATXN3* oraz *HTT*. Podczas analizy bioinformatycznej wyłoniłam białka, wpływające na wzór naprawy sekwencji powtarzającej się. Nowa wiedza uzyskana w ramach tych badań pozwoli dokładnie zgłębić mechanizmy naprawy DNA oraz być może stworzy możliwość sterowania nimi, co mogłoby stanowić jedno z potencjalnych podejść terapeutycznych w leczeniu chorób wywoływanych ekspansją powtórzeń trójnukleotydowych.

LITERATURA

[1] Ross, C. A. (2002). Polyglutamine pathogenesis: emergence of unifying mechanisms for Huntington's disease and related disorders. *Neuron*, *35*(5), 819-822

[2] Margulis, B. A., Vigont, V., Lazarev, V. F., Kaznacheyeva, E. V., & Guzhova, I. V. (2013). Pharmacological protein targets in polyglutamine diseases: mutant polypeptides and their interactors. *FEBS letters*, *587*(13), 1997-2007.

[3] Fan, H. C., Ho, L. I., Chi, C. S., Chen, S. J., Peng, G. S., Chan, T. M., ... & Harn, H. J. (2014). Polyglutamine (PolyQ) diseases: genetics to treatments. *Cell transplantation*, *23*(4-5), 441-458.

[4] Stevanin, G., & Brice, A. (2008). Spinocerebellar ataxia 17 (SCA17) and Huntington's disease-like 4 (HDL4). *The Cerebellum*, *7*(2), 170-178.

[5] Finsterer, J. (2010). Perspectives of Kennedy's disease. *Journal of the neurological sciences*, *298*(1-2), 1-10.

[6] Dion, V. (2014). Tissue specificity in DNA repair: lessons from trinucleotide repeat instability. *Trends in genetics*, *30*(6), 220-229.

[7] Stoyas, C. A., & La Spada, A. R. (2018). The CAG–polyglutamine repeat diseases: a clinical, molecular, genetic, and pathophysiologic nosology. *Handbook of clinical neurology*, *147*, 143-170.

[8] Mirkin, S. M. (2007). Expandable DNA repeats and human disease. *Nature*, *447*(7147), 932-940.

[9] Neil, A. J., Kim, J. C., & Mirkin, S. M. (2017). Precarious maintenance of simple DNA repeats in eukaryotes. *BioEssays*, *39*(9), 1700077.

[10] Solans, A., Zambrano, A., Rodríguez, M., & Barrientos, A. (2006). Cytotoxicity of a mutant huntingtin fragment in yeast involves early alterations in mitochondrial OXPHOS complexes II and III. *Human molecular genetics*, *15*(20), 3063-3081.

[11] Bennett, E. J., Shaler, T. A., Woodman, B., Ryu, K. Y., Zaitseva, T. S., Becker, C. H., ... & Kopito, R. R. (2007). Global changes to the ubiquitin system in Huntington's disease. *Nature*, *448*(7154), 704-708.

[12] Chafekar, S. M., & Duennwald, M. L. (2012). Impaired heat shock response in cells expressing full-length polyglutamine-expanded huntingtin. *PloS one*, *7*(5), e37929

[13] Fan, H. C., Ho, L. I., Chi, C. S., Chen, S. J., Peng, G. S., Chan, T. M., ... & Harn, H. J. (2014). Polyglutamine (PolyQ) diseases: genetics to treatments. *Cell transplantation*, *23*(4-5), 441-458.

[14] Gatchel, J. R., & Zoghbi, H. Y. (2005). Diseases of unstable repeat expansion: mechanisms and common principles. *Nature Reviews Genetics*, *6*(10), 743-755.

[15] Buijsen, R. A., Toonen, L. J., Gardiner, S. L., & van Roon-Mom, W. (2019). Genetics, mechanisms, and therapeutic progress in polyglutamine spinocerebellar ataxias. *Neurotherapeutics*, *16*(2), 263-286.

[16] Wild, E. J., & Tabrizi, S. J. (2017). Therapies targeting DNA and RNA in Huntington's disease. *The Lancet Neurology*, *16*(10), 837-847.

[17] Ashizawa, T., Öz, G., & Paulson, H. L. (2018). Spinocerebellar ataxias: prospects and challenges for therapy development. *Nature Reviews Neurology*, *14*(10), 590-605.

[18] Keiser, M. S., Kordasiewicz, H. B., & McBride, J. L. (2016). Gene suppression strategies for dominantly inherited neurodegenerative diseases: lessons from Huntington's disease and spinocerebellar ataxia. *Human molecular genetics*, *25*(R1), R53-R64.

[19] Mittelman, D., Moye, C., Morton, J., Sykoudis, K., Lin, Y., Carroll, D., & Wilson, J. H. (2009). Zinc-finger directed double-strand breaks within CAG repeat tracts promote repeat instability in human cells. *Proceedings of the National Academy of Sciences*, *106*(24), 9607-9612.

[20] Fink, K. D., Deng, P., Gutierrez, J., Anderson, J. S., Torrest, A., Komarla, A., ... & Nolta, J. A. (2016). Allele-specific reduction of the mutant huntingtin allele using transcription activator-like effectors in human huntington's disease fibroblasts. *Cell transplantation*, *25*(4), 677-686.

[21] Ouyang, S., Xie, Y., Xiong, Z., Yang, Y., Xian, Y., Ou, Z., ... & Sun, X. (2018). CRISPR/Cas9-targeted deletion of polyglutamine in spinocerebellar ataxia type 3-derived induced pluripotent stem cells. *Stem cells and development*, *27*(11), 756-770.

[22] Shin, J. W., Kim, K. H., Chao, M. J., Atwal, R. S., Gillis, T., MacDonald, M. E., ... & Lee, J. M. (2016). Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9. *Human molecular genetics*, *25*(20), 4566-4576.

[23] Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., ... & Koonin, E. V. (2011). Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology*, *9*(6), 467-477.

[24] 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*(6096), 816-821.

[25] Ran, F. A. F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, *8*(11), 2281-2308.

[26] Kosicki, M., Tomberg, K., & Bradley, A. (2018). Repair of double-strand breaks induced by

CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature biotechnology*, *36*(8), 765-771.

[27] Vakulskas, C. A., & Behlke, M. A. (2019). Evaluation and reduction of CRISPR off-target cleavage events. *Nucleic acid therapeutics*, *29*(4), 167-174.

[28] Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., ... & Zhang, F. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, *163*(3), 759-771.

[29] Kleinstiver, B. P., Tsai, S. Q., Prew, M. S., Nguyen, N. T., Welch, M. M., Lopez, J. M., ... & Joung, J. K. (2016). Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nature biotechnology*, *34*(8), 869-874.

[30] Swarts, D. C., & Jinek, M. (2018). Cas9 versus Cas12a/Cpf1: Structure–function comparisons and implications for genome editing. *Wiley Interdisciplinary Reviews: RNA*, *9*(5), e1481.

[31] Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., ... & Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, *154*(6), 1380-1389.

[32] Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., ... & Skarnes, W. C. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nature methods*, *11*(4), 399-402.

[33] Cullot, G., Boutin, J., Toutain, J., Prat, F., Pennamen, P., Rooryck, C., ... & Bedel, A. (2019). CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nature communications*, *10*(1), 1-14.

[34] Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., ... & Qi, L. S. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, *154*(2), 442-451.

[35] Thakore, P. I., D'ippolito, A. M., Song, L., Safi, A., Shivakumar, N. K., Kabadi, A. M., ... & Gersbach, C. A. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nature methods*, *12*(12), 1143-1149.

[36] Zhang, J. P., Li, X. L., Li, G. H., Chen, W., Arakaki, C., Botimer, G. D., ... & Zhang, X. B. (2017). Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome biology*, *18*(1), 1-18.

[37] Yoshimi, K., Oka, Y., Miyasaka, Y., Kotani, Y., Yasumura, M., Uno, Y., ... & Mashimo, T. (2021). Combi-CRISPR: combination of NHEJ and HDR provides efficient and precise plasmid-based knock-ins in mice and rats. *Human genetics*, *140*(2), 277-287.

[38] Ma, H., Tu, L. C., Naseri, A., Huisman, M., Zhang, S., Grunwald, D., & Pederson, T. (2016).

Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nature biotechnology*, *34*(5), 528-530.

[39] Hong, Y., Lu, G., Duan, J., Liu, W., & Zhang, Y. (2018). Comparison and optimization of CRISPR/dCas9/gRNA genome-labeling systems for live cell imaging. *Genome biology*, *19*(1), 1-10.

[40] Horlbeck, M. A., Gilbert, L. A., Villalta, J. E., Adamson, B., Pak, R. A., Chen, Y., ... & Weissman, J. S. (2016). Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *elife*, *5*, e19760.

[41] Tian, R., Gachechiladze, M. A., Ludwig, C. H., Laurie, M. T., Hong, J. Y., Nathaniel, D., ... & Kampmann, M. (2019). CRISPR interference-based platform for multimodal genetic screens in human iPSC-derived neurons. *Neuron*, *104*(2), 239-255.

[42] Savić, N., & Schwank, G. (2016). Advances in therapeutic CRISPR/Cas9 genome editing. *Translational Research*, *168*, 15-21.

[43] Chen, M., Mao, A., Xu, M., Weng, Q., Mao, J., & Ji, J. (2019). CRISPR-Cas9 for cancer therapy: Opportunities and challenges. *Cancer letters*, *447*, 48-55.

[44] Heman-Ackah, S. M., Bassett, A. R., & Wood, M. J. A. (2016). Precision modulation of neurodegenerative disease-related gene expression in human iPSC-derived neurons. *Scientific reports*, *6*(1), 1-12.

[45] Wu, J., Tang, Y., & Zhang, C. L. (2019). Targeting N-terminal Huntingtin with a dual-sgRNA strategy by CRISPR/Cas9. *BioMed Research International*, *2019*.

[46] Monteys, A. M., Ebanks, S. A., Keiser, M. S., & Davidson, B. L. (2017). CRISPR/Cas9 editing of the mutant huntingtin allele in vitro and in vivo. *Molecular Therapy*, *25*(1), 12-23.

[47] Kolli, N., Lu, M., Maiti, P., Rossignol, J., & Dunbar, G. L. (2017). CRISPR-Cas9 mediated gene-silencing of the mutant huntingtin gene in an in vitro model of Huntington's disease. *International journal of molecular sciences*, *18*(4), 754.

[48] He, L., Wang, S., Peng, L., Zhao, H., Li, S., Han, X., ... & Jiang, H. (2021). CRISPR/Cas9 mediated gene correction ameliorates abnormal phenotypes in spinocerebellar ataxia type 3 patient-derived induced pluripotent stem cells. *Translational psychiatry*, *11*(1), 1-13.

[49] Yang, S., Chang, R., Yang, H., Zhao, T., Hong, Y., Kong, H. E., ... & Li, X. J. (2017). CRISPR/Cas9-mediated gene editing ameliorates neurotoxicity in mouse model of Huntington's disease. *The Journal of clinical investigation*, *127*(7), 2719-2724.

[50] Vouillot, L., Thélie, A., & Pollet, N. (2015). Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3: Genes, Genomes, Genetics*, *5*(3), 407-415.

[51] Zischewski, J., Fischer, R., & Bortesi, L. (2017). Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnology advances*, *35*(1), 95-104.

[52] Yang, Z., Steentoft, C., Hauge, C., Hansen, L., Thomsen, A. L., Niola, F., ... & Bennett, E. P. (2015). Fast and sensitive detection of indels induced by precise gene targeting. *Nucleic acids research*, *43*(9), e59-e59.

[53] Brinkman, E. K., Chen, T., Amendola, M., & van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic acids research*, *42*(22), e168-e168.

[54] Brinkman, E. K., Kousholt, A. N., Harmsen, T., Leemans, C., Chen, T., Jonkers, J., & van Steensel, B. (2018). Easy quantification of template-directed CRISPR/Cas9 editing. *Nucleic acids research*, *46*(10), e58-e58.

[55] Findlay, S. D., Vincent, K. M., Berman, J. R., & Postovit, L. M. (2016). A digital PCR-based method for efficient and highly specific screening of genome edited cells. *PloS one*, *11*(4), e0153901.

[56] Hendel, A., Kildebeck, E. J., Fine, E. J., Clark, J. T., Punjya, N., Sebastiano, V., ... & Porteus, M. H. (2014). Quantifying genome-editing outcomes at endogenous loci with SMRT sequencing. *Cell reports*, *7*(1), 293-305.

[57] Ebbert, M. T., Farrugia, S. L., Sens, J. P., Jansen-West, K., Gendron, T. F., Prudencio, M., ... & Fryer, J. D. (2018). Long-read sequencing across the C9orf72 'GGGGCC'repeat expansion: implications for clinical use and genetic discovery efforts in human disease. *Molecular neurodegeneration*, *13*(1), 1-17.

[58] Höijer, I., Johansson, J., Gudmundsson, S., Chin, C. S., Bunikis, I., Häggqvist, S., ... & Ameur, A. (2020). Amplification-free long-read sequencing reveals unforeseen CRISPR-Cas9 off-target activity. *Genome biology*, *21*(1), 1-19.

[59] 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*(10), 1473-1475.

[60] Concordet, J. P., & Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic acids research*, *46*(W1), W242-W245.

[61] Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., & Kim, J. S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome research*, *24*(1), 132-141.

[62] Smith, C., Gore, A., Yan, W., Abalde-Atristain, L., Li, Z., He, C., ... & Ye, Z. (2014). Whole-

genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell stem cell*, *15*(1), 12-13.

[63] Iyer, V., Shen, B., Zhang, W., Hodgkins, A., Keane, T., Huang, X., & Skarnes, W. C. (2015). Off-target mutations are rare in Cas9-modified mice. *Nature methods*, *12*(6), 479-479.

[64] Crosetto, N., Mitra, A., Silva, M. J., Bienko, M., Dojer, N., Wang, Q., ... & Dikic, I. (2013). Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nature methods*, *10*(4), 361-365.

[65] Saudou, F., & Humbert, S. (2016). The biology of huntingtin. *Neuron*, *89*(5), 910-926.

[66] Morozova, K. N., Suldina, L. A., Malankhanova, T. B., Grigor'eva, E. V., Zakian, S. M., Kiseleva, E., & Malakhova, A. A. (2018). Introducing an expanded CAG tract into the huntingtin gene causes a wide spectrum of ultrastructural defects in cultured human cells. *PLoS One*, *13*(10), e0204735.

[67] Xu, X., Tay, Y., Sim, B., Yoon, S. I., Huang, Y., Ooi, J., ... & Pouladi, M. A. (2017). Reversal of phenotypic abnormalities by CRISPR/Cas9-mediated gene correction in Huntington disease patient-derived induced pluripotent stem cells. *Stem cell reports*, *8*(3), 619-633.

[68] An, M. C., O'Brien, R. N., Zhang, N., Patra, B. N., De La Cruz, M., Ray, A., & Ellerby, L. M. (2014). Polyglutamine disease modeling: epitope based screen for homologous recombination using CRISPR/Cas9 system. *PLoS currents*, *6*.

[69] Marthaler, A. G., Schmid, B., Tubsuwan, A., Poulsen, U. B., Engelbrecht, A. F., Mau-Holzmann, U. A., ... & Holst, B. (2016). Generation of an isogenic, gene-corrected control cell line of the spinocerebellar ataxia type 2 patient-derived iPSC line H196. *Stem cell research*, *16*(1), 162-165.

[70] Marthaler, A. G., Schmid, B., Tubsuwan, A., Poulsen, U. B., Engelbrecht, A. F., Mau-Holzmann, U. A., ... & Holst, B. (2016). Generation of an isogenic, gene-corrected control cell line of the spinocerebellar ataxia type 2 patient-derived iPSC line H196. *Stem cell research*, *16*(1), 162-165.

[71] Marthaler, A. G., Tubsuwan, A., Schmid, B., Poulsen, U. B., Engelbrecht, A. F., Mau-Holzmann, U. A., ... & Holst, B. (2016). Generation of an isogenic, gene-corrected control cell line of the spinocerebellar ataxia type 2 patient-derived iPSC line H266. *Stem cell research*, *16*(1), 202-205.

[72] Zuo, E., Huo, X., Yao, X., Hu, X., Sun, Y., Yin, J., ... & Yang, H. (2017). CRISPR/Cas9-mediated targeted chromosome elimination. *Genome biology*, *18*(1), 1-18.

[73] Marcinkowska, M., Wong, K. K., Kwiatkowski, D. J., & Kozlowski, P. (2010). Design and generation of MLPA probe sets for combined copy number and small-mutation analysis of human genes: EGFR as an example. *TheScientificWorldJOURNAL*, *10*, 2003-2018.

[74] R. R. Brinkman, M. M. Mezei, J. Theilmann, E. Almqvist, and M. R. Hayden, “The likelihood of being affected with huntington disease by a particular age, for a specific CAG size,” *Am. J. Hum. Genet.*, vol. 60, no. 5, pp. 1202–1210, 1997.

[75] Quarrell, O. W., Nance, M. A., Nopoulos, P., Paulsen, J. S., Smith, J. A., & Squitieri, F. (2013). Managing juvenile Huntington’s disease. *Neurodegenerative disease management*, *3*(3), 267-276.

[76] Harper, S. Q., Staber, P. D., He, X., Eliason, S. L., Martins, I. H., Mao, Q., ... & Davidson, B. L. (2005). RNA interference improves motor and neuropathological abnormalities in a Huntington's disease mouse model. *Proceedings of the National Academy of Sciences*, *102*(16), 5820-5825.

[77] Boudreau, R. L., McBride, J. L., Martins, I., Shen, S., Xing, Y., Carter, B. J., & Davidson, B. L. (2009). Nonallele-specific silencing of mutant and wild-type huntingtin demonstrates therapeutic efficacy in Huntington's disease mice. *Molecular Therapy*, *17*(6), 1053-1063.

[78] Popp, M. W., & Maquat, L. E. (2016). Leveraging rules of nonsense-mediated mRNA decay for genome engineering and personalized medicine. *Cell*, *165*(6), 1319-1322.

[79] Sun, J., & Roy, S. (2021). Gene-based therapies for neurodegenerative diseases. *Nature neuroscience*, *24*(3), 297-311

[80] Kolli, Nivya, et al. "CRISPR-Cas9 mediated gene-silencing of the mutant huntingtin gene in an in vitro model of Huntington’s disease." *International journal of molecular sciences* 18.4 (2017): 754.

[81] Fiszer, A., & Krzyzosiak, W. J. (2013). RNA toxicity in polyglutamine disorders: concepts, models, and progress of research. *Journal of molecular medicine*, *91*(6), 683-691.

[82] Nalavade, R., Griesche, N., Ryan, D. P., Hildebrand, S., & Krauss, S. (2013). Mechanisms of RNA-induced toxicity in CAG repeat disorders. *Cell death & disease*, *4*(8), e752-e752.

[83] Fiszer, A., Olejniczak, M., Galka-Marciniak, P., Mykowska, A., & Krzyzosiak, W. J. (2013). Self-duplexing CUG repeats selectively inhibit mutant huntingtin expression. *Nucleic acids research*, *41*(22), 10426-10437.

[84] Cinesi, C., Aeschbach, L., Yang, B., & Dion, V. (2016). Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase. *Nature communications*, *7*(1), 1-10.

OŚWIADCZENIA DOKTORANTA

Oświadczenia doktoranta dotyczące jego udziału w powstaniu
prac naukowych stanowiących rozprawę doktorską

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 01.03.2022

Oświadczenie doktoranta dotyczące jego udziału w powstawaniu pracy naukowej wchodzącej w skład rozprawy doktorskiej

Dabrowska, M., Juzwa, W., Krzyzosiak, W. J., & Olejniczak, M. (2018). **Precise excision of the CAG tract from the huntingtin gene by Cas9 nickases.** *Frontiers in Neuroscience*, *12*, 75.

Oświadczam, że mój udział w tworzeniu tej pracy polegał na: przeprowadzeniu wszystkich eksperymentów oraz pomocy w opracowaniu rycin i materiałów suplementarnych. Eksperymenty były związane z projektowaniem gRNA oraz konstruowaniem plazmidów z ekspresją białek Cas9, nikaza Cas9 i gRNA. Ponadto, za pomocą różnych metod wprowadzałam komponenty tych systemów do komórek HEK 293T oraz fibroblastów HD z różną długością sekwencji CAG. Samodzielnie wykonałam analizy, które miały na celu określenie efektywności edycji w regionie sekwencji CAG oraz wykrycie potencjalnych efektów niespecyficznych. Oprócz tego badałam efekt usunięcia sekwencji powtarzającej się na poziom transkryptu HTT oraz białka.

Magdalena Dąbrowska
.....

Podpis

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 01.03.2022

Oświadczenie doktoranta dotyczące jego udziału w powstawaniu pracy naukowej wchodzącej w skład rozprawy doktorskiej

Dabrowska, M., & Olejniczak, M. (2020). **Gene therapy for Huntington's disease using targeted endonucleases.** In *Trinucleotide Repeats* (pp. 269-284). Humana, New York, NY.

Oświadczam, że mój udział w powstawaniu tej pracy polegał na: opracowaniu protokołu do inaktywacji huntingtyny poprzez użycie dwóch cząsteczek gRNA i nikazy Cas9. Ponadto wykonałam rysunki i brałam udział w pisaniu manuskryptu.

Magdalena Dąbrowska
Podpis

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 01.03.2022

Oświadczenie doktoranta dotyczące jego udziału w powstawaniu pracy naukowej wchodzącej w skład rozprawy doktorskiej

Dabrowska, M., Czubak, K., Juzwa, W., Krzyzosiak, W. J., Olejniczak, M., & Kozlowski, P. (2018). **qEva-CRISPR: a method for quantitative evaluation of CRISPR/Cas-mediated genome editing in target and off-target sites.** *Nucleic Acids Research*, *46*(17), e101-e101.

Oświadczam, że mój udział w powstawaniu tej pracy polegał na: projektowaniu i wykonaniu eksperymentów oraz na pomocy w pisaniu manuskryptu, przygotowaniu rycin i materiałów suplementarnych. Wykonałam wszystkie eksperymenty edycji genomu. Zaprojektowałam komponenty systemu CRISPR-Cas9 oraz skonstruowałam plazmidy z ich ekspresją. Ponadto zoptymalizowałam metody dostarczania komponentów systemu CRISPR-Cas9 oraz metody izolacji materiału genetycznego dla czterech typów komórek. Do moich zadań należała również synteza gRNA celujących w gen *HTT* przy pomocy transkrypcji *in vitro*. Ponadto uczestniczyłam w koncepcji tworzenia metody i projektowaniu sond.

Podpis

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 01.03.2022

Oświadczenie doktoranta dotyczące jego udziału w powstawaniu pracy naukowej wchodzącej w skład rozprawy doktorskiej

Dabrowska, M., Ciolak, A., Kozlowska, E., Fiszer, A., & Olejniczak, M. (2020). **Generation of new isogenic models of Huntington's disease using CRISPR-Cas9 technology**. *International Journal of Molecular Sciences*, *21*(5), 1854.

Oświadczam, że mój udział w powstawaniu tej pracy polegał na: zaprojektowaniu eksperymentów, pomocy w pisaniu manuskryptu i przygotowania rycin oraz materiałów suplementarnych. Opracowałam strategie otrzymywania linii komórkowych z wydłużoną i skróconą sekwencja CAG w genie *HTT*. Zaprojektowałam i stworzyłam matryce donorowe. Wykonałam eksperymenty związane z dostarczaniem komponentów systemu CRISPR-Cas9 i CRISPR-Cas9n do iPSCs i komórek HEK 293T. Mój udział w tworzeniu tej publikacji polegał również na charakterystyce otrzymanych linii HEK 293T z wydłużoną sekwencją CAG w genie *HTT* oraz zbadaniu zaburzeń składania mRNA, jakim jest powstawanie transkryptu HTT z intronem 1. Ponadto, na wyprowadzonych liniach przetestowałam potencjał wyciszający dwóch cząsteczek siRNA.

Podpis

OŚWIADCZENIA AUTORÓW KORESPONDENCYJNYCH

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 01.03.2022

dr hab. Marta Olejniczak
Zakład Inżynierii Genomowej
Instytut Chemii Bioorganicznej PAN

Oświadczenie autora korespondencyjnego

o udziale doktoranta w powstaniu artykułu naukowego

Niniejszym potwierdzam, że w publikacji:

Dabrowska, M., Juzwa, W., Krzyzosiak, W. J., & Olejniczak, M. (2018). **Precise excision of the CAG tract from the huntingtin gene by Cas9 nickases.** *Frontiers in Neuroscience*, *12*, 75.

jestem autorem korespondencyjnym oraz, że wkład autorski w ww. publikację mgr Magdaleny Dąbrowskiej polegał na przeprowadzeniu wszystkich eksperymentów oraz pomocy w opracowaniu rycin i materiałów suplementarnych.
Doktorantka zaprojektowała gRNA oraz skonstruowała wektory plazmidowe do ekspresji Cas9, Cas9n i gRNA. Ponadto, za pomocą różnych metod, wprowadzała komponenty tych systemów do komórek HEK 293T oraz fibroblastów HD z różną długością sekwencji CAG. Mgr Magdalena Dąbrowska wykonała również analizy, które miały na celu określenie efektywności edycji w regionie sekwencji CAG oraz wykrycie potencjalnych efektów niespecyficznych, jak również badała efekt usunięcia sekwencji powtarzającej się na poziom transkryptu HTT oraz białka.

Mój udział w tworzeniu tej publikacji polegał na opracowaniu koncepcji badań, interpretacji danych uzyskanych w wyniku przeprowadzonych analiz, napisaniu manuskryptu i opracowaniu materiałów suplementarnych. Ponadto nadzorowałam pracę eksperymentalną Doktorantki i zdobyłam fundusze na te badania.

Marta Olejniczak

Podpis autora korespondencyjnego

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 01.03.2022

dr hab. Marta Olejniczak
Zakład Inżynierii Genomowej
Instytut Chemii Bioorganicznej PAN

Oświadczenie autora korespondencyjnego

o udziale doktoranta w powstaniu artykułu naukowego

Niniejszym potwierdzam, że w publikacji:

Dabrowska, M., & Olejniczak, M. (2020). **Gene therapy for Huntington's disease using targeted endonucleases.** In *Trinucleotide Repeats* (pp. 269-284). Humana, New York, NY.

jestem autorem korespondencyjnym oraz, że wkład autorski w ww. publikację mgr Magdaleny Dąbrowskiej polegał na opracowaniu protokołu do inaktywacji huntingtyny przy użyciu systemu CRISPR-Cas9n. Ponadto Doktorantka wykonała rysunki i brała udział w pisaniu manuskryptu.

Mój udział w tworzeniu tej publikacji polegał na opracowaniu wspólnie z mgr Magdaleną Dąbrowską koncepcji protokołu, napisaniu manuskryptu i nadzorowaniu pracy Doktorantki.

Marta Olejniczak

Podpis autora korespondencyjnego

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 01.03.2022

dr hab. Marta Olejniczak
Zakład Inżynierii Genomowej
Instytut Chemii Bioorganicznej PAN

Oświadczenie autora korespondencyjnego

o udziale doktoranta w powstaniu artykułu naukowego

Niniejszym potwierdzam, że w publikacji:

Dabrowska, M., Czubak, K., Juzwa, W., Krzyzosiak, W. J., Olejniczak, M., & Kozlowski, P. (2018). **qEva-CRISPR: a method for quantitative evaluation of CRISPR/Cas-mediated genome editing in target and off-target sites.** *Nucleic Acids Research*, *46*(17), e101-e101.

jestem autorem korespondencyjnym oraz, że wkład autorski w ww. publikację mgr Magdaleny Dąbrowskiej polegał na projektowaniu i wykonaniu eksperymentów oraz na pomocy w pisaniu manuskryptu, przygotowaniu rycin i materiałów suplementarnych.
Doktorantka zaprojektowała i wykonała wszystkie eksperymenty edycji genomu opisane w publikacji. Zoptymalizowała metody dostarczania komponentów systemu CRISPR-Cas9 oraz metody izolacji materiału genetycznego dla czterech typów komórek stosowanych w badaniach. Do jej zadań należała również synteza gRNA celujących w gen *HTT* przy pomocy transkrypcji *in vitro*. Doktorantka uczestniczyła również w tworzeniu koncepcji metody qEva-CRISPR i projektowaniu sond.

Mój udział w tworzeniu tej publikacji polegał na uczestniczeniu w opracowaniu koncepcji badań, interpretacji danych uzyskanych w wyniku przeprowadzonych analiz, napisaniu manuskryptu i opracowaniu materiałów suplementarnych. Ponadto nadzorowałam pracę eksperymentalną Doktorantki i zdobyłam fundusze na te badania.

Marta Olejniczak

Podpis autora korespondencyjnego

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 01.03.2022

Prof. dr hab. Piotr Kozłowski
Zakład Genetyki Molekularnej (kierownik)
Instytut Chemii Bioorganicznej PAN

Oświadczenie autora korespondencyjnego

o udziale doktoranta w powstaniu artykułu naukowego

Niniejszym potwierdzam, że razem z dr hab. Martą Olejniczak jestem autorem korespondencyjnym poniżej wymienionej publikacji:

Dabrowska, M., Czubak, K., Juzwa, W., Krzyzosiak, W. J., Olejniczak, M., & Kozlowski, P. (2018). **qEva-CRISPR: a method for quantitative evaluation of CRISPR/Cas-mediated genome editing in target and off-target sites.** *Nucleic Acids Research*, *46*(17), e101-e101.

Badania opisane w publikacji zostały wykonane w ramach współpracy między moją grupą (Zakład Genetyki Molekularnej) i grupą dr hab. Marty Olejniczak (Zakład Inżynierii Genomowej). W ramach przeprowadzonych badań mgr Magdalena Dąbrowska wykonała wszystkie eksperymenty i analizy związane z edycją genomu.

Doktorantka zaprojektowała i wykonała wszystkie eksperymenty edycji genomu opisane w publikacji. Zoptymalizowała metody dostarczania komponentów systemu CRISPR-Cas9 oraz metody izolacji materiału genetycznego dla czterech typów komórek stosowanych w badaniach. Do jej zadań należała również synteza gRNA celujących w gen *HTT* przy pomocy transkrypcji *in vitro*. Doktorantka uczestniczyła również w tworzeniu koncepcji metody qEva-CRISPR i projektowaniu sond.
Ponadto mgr Magdalena Dąbrowska brała udział w dyskutowaniu wyników, opracowaniu kolejnych kroków analiz, podsumowaniu wyników, oraz pomagała w przygotowaniu manuskryptu, rycin i materiałów suplementarnych.

Mój udział w tworzeniu tej publikacji (wspólnie z dr hab. Martą Olejniczak) polegał na opracowaniu koncepcji, koordynacji badań oraz nadzorze nad pracą doktorantów i zdobyciu środków na finansowanie badań. Brałem również udział w interpretacji danych uzyskanych w wyniku przeprowadzonych analiz i przygotowaniu manuskryptu.

Piotr Kozłowski

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 01.03.2022

dr hab. Marta Olejniczak
Zakład Inżynierii Genomowej
Instytut Chemii Bioorganicznej PAN

Oświadczenie autora korespondencyjnego

o udziale doktoranta w powstaniu artykułu naukowego

Niniejszym potwierdzam, że w publikacji:

Dabrowska, M., Ciolak, A., Kozlowska, E., Fiszer, A., & Olejniczak, M. (2020). **Generation of new isogenic models of Huntington's disease using CRISPR-Cas9 technology**. *International Journal of Molecular Sciences*, *21*(5), 1854.

jestem autorem korespondencyjnym oraz, że wkład autorski w ww. publikację mgr Magdaleny Dąbrowskiej polegał na zaprojektowaniu i wykonaniu części eksperymentów, pomocy w pisaniu manuskryptu i przygotowaniu rycin oraz materiałów suplementarnych.
Doktorantka opracowała strategie otrzymywania linii komórkowych z wydłużoną i skróconą sekwencją CAG w genie *HTT*. Ponadto zaprojektowała i stworzyła matryce donorowe oraz wykonała eksperymenty związane z edycją genu *HTT* przy pomocy systemu CRISPR-Cas9 i CRISPR-Cas9n w iPSCs i komórkach HEK 293T. Doktorantka scharakteryzowała otrzymane linie HEK 293T z wydłużoną sekwencją CAG w genie *HTT.* W tych komórkach zbadała proces nieprawidłowego składania mRNA, poprzez analizę poziomu transkryptu HTT z intronem 1 oraz potwierdziła potencjał wyciszający dwóch cząsteczek siRNA.

Mój udział w tworzeniu tej publikacji polegał na opracowaniu wspólnie z mgr Magdaleną Dąbrowską koncepcji badań, interpretacji danych uzyskanych w wyniku przeprowadzonych analiz, napisaniu manuskryptu i opracowaniu materiałów suplementarnych. Ponadto nadzorowałam pracę eksperymentalną Doktorantki.

Marta Olejniczak
.....
Podpis autora korespondencyjnego

PUBLIKACJE STANOWIĄCE PODSTAWĘ PRACY DOKTORSKIEJ

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 26 February 2018
doi: 10.3389/fnins.2018.00075

Precise Excision of the CAG Tract from the Huntingtin Gene by Cas9 Nickases

Magdalena Dabrowska[1], *Wojciech Juzwa*[2], *Wlodzimierz J. Krzyzosiak*[3] *and Marta Olejniczak*[1]*

[1] Department of Genome Engineering, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, [2] Department of Biotechnology and Food Microbiology, Poznan University of Life Sciences, Poznan, Poland, [3] Department of Molecular Biomedicine, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

Huntington's disease (HD) is a progressive autosomal dominant neurodegenerative disorder caused by the expansion of CAG repeats in the first exon of the huntingtin gene (*HTT*). The accumulation of polyglutamine-rich huntingtin proteins affects various cellular functions and causes selective degeneration of neurons in the striatum. Therapeutic strategies used to date to silence the expression of mutant *HTT* include antisense oligonucleotides, RNA interference-based approaches and, recently, genome editing with the CRISPR/Cas9 system. Here, we demonstrate that the CAG repeat tract can be precisely excised from the *HTT* gene with the use of the paired Cas9 nickase strategy. As a model, we used HD patient-derived fibroblasts with varied numbers of CAG repeats. The repeat excision inactivated the *HTT* gene and abrogated huntingtin synthesis in a CAG repeat length-independent manner. Because Cas9 nickases are known to be safe and specific, our approach provides an attractive treatment tool for HD that can be extended to other polyQ disorders.

Keywords: genome editing, CRISPR/Cas9, neurodegenerative diseases, repeat expansion, engineered nucleases, Huntington's disease, nonsense-mediated decay

OPEN ACCESS

Edited by:
Sandro Alves,
Brainvectis Therapeutics, France

Reviewed by:
Nicole Déglon,
Centre Hospitalier Universitaire Vaudois (CHUV), Switzerland
Panchanan Maiti,
University of California, Los Angeles, United States

***Correspondence:**
Marta Olejniczak
marta.olejniczak@ibch.poznan.pl

Specialty section:
This article was submitted to Neurodegeneration, a section of the journal Frontiers in Neuroscience

Received: *28 November 2017*
Accepted: *29 January 2018*
Published: *26 February 2018*

Citation:
Dabrowska M, Juzwa W, Krzyzosiak WJ and Olejniczak M (2018) Precise Excision of the CAG Tract from the Huntingtin Gene by Cas9 Nickases. Front. Neurosci. 12:75. doi: 10.3389/fnins.2018.00075

INTRODUCTION

Expansions of short tandem repeat sequences in functionally unrelated genes are causative factors of numerous human hereditary neurological diseases. Currently, there are nine known neurodegenerative disorders caused by the expansion of CAG repeats within the coding regions of associated genes. These disorders include Huntington's disease (HD) (Bates et al., 2015); spinocerebellar ataxia types 1, 2, 3, 6, 7, and 17 (SCA) (Paulson et al., 2017); spinal-bulbar muscular atrophy (SBMA) (Spada et al., 1991); and dentatorubral-pallidoluysian atrophy (DRPLA) (Koide et al., 1994). A positive correlation exists between the size of the expansion and the severity of symptoms, which usually appear during the 4th-5th decade of life and lead to patient's death (Duyao et al., 1993).

HD is caused by the expansion of CAG repeats in exon 1 of the *HTT* gene, which encodes huntingtin (HTT), a large protein of more than 3,000 amino acids (Saudou and Humbert, 2016). Expanded polyglutamine (polyQ) protein may form intracellular aggregates and affects numerous cellular activities inducing pathogenesis *via* a gain of toxic function. Despite many years of research on an effective treatment method, HD and other polyQ diseases are incurable, and only their symptoms can be controlled. Several different strategies have already been employed in cellular and animal models of polyQ diseases to achieve the desired therapeutic effects

(Wild and Tabrizi, 2017). These strategies include the silencing of both HTT alleles in a non-allele-selective strategy and the targeting of single-nucleotide polymorphisms (SNPs) linked to repeat expansions. The repeat region itself may be targeted in an allele selective and non-selective manner (Fiszer et al., 2012; Keiser et al., 2016; Esteves et al., 2017). RNA interference and antisense oligonucleotide technologies, which have been used for many years in experimental therapy for polyQ diseases, are currently complemented with genome editing systems such as the CRISPR/Cas9 (Shin et al., 2016; Kolli et al., 2017; Merienne et al., 2017; Monteys et al., 2017; Yang et al., 2017).

Zinc finger nucleases (ZFNs) and transcription activator-like effector-based nucleases (TALENs) were the first tools that provided proof of principle for the idea of targeted inactivation of the expanded CAG repeats at a disease *loci* (Mittelman et al., 2009; Richard et al., 2014). In one of the first studies, preceding the CRISPR/Cas9 technology development, Isalan group used zinc finger proteins (ZFPs) to selectively bind and repress expanded CAG repeats in the R6/2 mouse model of HD (Garriga-Canut et al., 2012). In other approach expanded CAG repeat tracts were replaced with a normal CAG length by inducing homologous recombination in induced pluripotent stem cells (iPSCs) derived from HD patient fibroblasts (An et al., 2012). The efficiency of homologous recombination was further increased by using CRISPR/Cas9 (An et al., 2014).

The CRISPR-Cas9 system uses a small guide RNA (sgRNA) containing a 20 nt sequence complementary to the target DNA and Cas9 nuclease for site-specific cleavage of a genomic target containing a protospacer-adjacent motif (PAM) (Jinek et al., 2012). Double-strand breaks (DSBs) are repaired mainly by error-prone non-homologous end joining (NHEJ), resulting in mutations that may cause frame-shifts in open reading frames, premature translation termination and transcript degradation by nonsense-mediated decay (NMD). To increase specificity and reduce off-targeting, one of two cleavage domains in the Cas9 protein was mutated to act as a nickase (Cas9n) (Cho et al., 2014; Trevino and Zhang, 2014). Nickases generate single strand breaks (SSBs) that are repaired with high fidelity. Paired sgRNA/Cas9 nickases targeted to the opposite DNA strands enable genome editing via homology-directed repair (HDR) and have been shown to reduce off-targeting by 5- to 1.500-fold compared to wild-type Cas9 (wt Cas9) (Ran et al., 2013; Cho et al., 2014). Therefore, the paired Cas9 nickase strategy can be useful in applications that require precise genome editing such as gene and cell therapy.

To date, the CRISPR/Cas9 system has been used to selectively inactivate mutant *HTT* genes by using PAM sites generated by SNP alleles (Shin et al., 2016; Monteys et al., 2017). Although this strategy is very promising, it requires a comprehensive analysis of the *HTT* gene haplotype structure. In addition, the non-allele selective approach has been used to inactivate the *HTT* gene by using a pair of sgRNAs flanking CAG repeats and wt Cas9 in a transgenic mouse model of HD (Yang et al., 2017). Non-allele selective supression of *HTT* gene expression was achieved also by using CRISPR interference strategy (CRISPRi) in HEK293T cells (Heman-Ackah et al., 2016). In this approach nuclease null, dead Cas9 (dCas9) and sgRNAs targeting *HTT* transcription start site were used.

In this study, we examined paired Cas9 nickase strategy to inactivate the *HTT* gene by targeting sequences directly flanking the CAG repeat tract. We demonstrate that precise excision of the CAG repeats from the *HTT* gene results in the abrogation of protein synthesis in all investigated fibroblast cell lines derived from HD patients. Importantly, we also show that this specific and safe strategy leads to preservation of repeat-deficient transcript level, suggesting that the transcript may escape from NMD pathway.

MATERIALS AND METHODS

Cell Culture and Transfection

Fibroblasts (GM04208, 21/44 CAG in the *HTT* gene; GM04281, 17/68 CAG in the *HTT* gene; GM09197, 21/151 CAG in the *HTT* gene) were obtained from the Coriell Cell Repositories (Camden, New Jersey, USA) and grown in minimal essential medium (Lonza; Basel, Switzerland) supplemented with 10% fetal bovine serum (Sigma-Aldrich; St. Louis, MO, USA), antibiotics (Sigma-Aldrich, A5955) and non-essential amino acids (Sigma-Aldrich, M7145). HEK293T cells (16/17 CAG in the *HTT* gene) were grown in Dulbecco's modified Eagle's medium (Lonza; Basel, Switzerland) supplemented with 10% fetal bovine serum (Sigma-Aldrich), antibiotics (Sigma-Aldrich) and L-glutamine (Sigma-Aldrich). HEK293T transfections were performed using calcium phosphate method with 10 μg of plasmid DNA for 3×10^5 cells (Jordan et al., 1996). Fibroblasts were electroporated with the NeonTM Transfection System (Invitrogen, Carlsbad, CA, USA). Briefly, 1×10^6 to 5×10^5 cells were harvested, resuspended in PBS and electroporated with 10 μg of plasmid DNA (5 μg of each plasmid from a HTT_sgRNA/Cas9n pair) in 100 μl tips using the following parameters: 1.350 V, 30 ms, 1 pulse. Fibroblasts were sorted by flow cytometry (BD Biosciences, BD FACS AriaIII) 48 h post electroporation and collected for genomic DNA, RNA and protein extraction.

Plasmids

Guide RNA sequences for the CRISPR/Cas9 system were designed as described in Ran et al. with the use of CRISPOR software (http://crispor.tefor.net/crispor.py; Haeussler et al., 2016). Briefly, the top and bottom strand of 20-nt guide RNA were synthesized (IBB, Warsaw), annealed and ligated into the pair of FastDigest BsmBI (Thermo Fisher Scientific, Waltham, MA, USA) cut plasmids, namely, pSpCas9(BB)-2A-GFP (PX458) (Addgene, Cambridge, MA, USA) and its nickase version (D10A nickase mutant; pSpCas9n(BB)-2A-GFP (PX461)) from *S. pyogenes* (Ran et al., 2013). Ligated products were transformed into chemically competent E. coli GT116 cells (InvivoGen, San Diego, CA, USA), and the cells were plated onto ampicillin selection plates (100 μg/mL ampicillin) and incubated at 37°C overnight. Plasmid DNA was isolated using the Gene JET Plasmid Miniprep kit (Thermo Scientific) and verified with Sanger sequencing. For larger scale plasmid preparations, the Qiagen Midi kit was used (Qiagen, Hilden, Germany). The sgRNA oligonucleotide sequences are presented in **Table S1**.

DNA Extraction and Analysis of Genome Editing Efficiency

Genomic DNA from the HD fibroblast and HEK293T cell lines was extracted using the Cells and Tissue DNA Isolation Kit (Norgen, Biotek Corp., Schmon Pkwy, ON, Canada) according to manufacturer's instructions and quantified using a spectrophotometer/fluorometer (DeNovix, Wilmington, DE, USA). For the T7E1 mismatch analysis, genomic DNA was amplified using Phusion High-Fidelity PCR Master Mix (Thermo Fisher) with primers HD1F and HD1R spanning CAG repeats in exon 1 of the *HTT* gene. The two-step PCR amplification program was used as follows: an initial denaturation at 98°C for 3 min; 12 cycles at 98°C for 15 s, 72°C for 15 s; 21 cycles at 98°C for 15 s, 62°C for 15 s, and 72°C for 15 s; and a final elongation at 72°C for 5 min. PCR products were purified using the GeneJET PCR Purification Kit (Thermo Fisher). 400 ng of the purified PCR product was used in an annealing reaction and enzymatic digestion with the T7E1 enzyme (New England Biolabs, Ipswich, MA, USA). Cleavage products were separated in 1.3% agarose gels and detected using G-BOX. Band intensities were analyzed with GelPro software (Media Cybernetics, Rockville, MD, USA). Indel occurrence was estimated with an analysis of signal loss from the main PCR products. Briefly, the main band intensities from HTT_sgRNA-treated samples were compared to the same bands from control samples treated with the empty plasmid. Genes selected for off-target analysis were PCR-amplified with specific primer pairs (**Table S1**) using Phusion High-Fidelity PCR Master Mix (Thermo Fisher) and the following program: an initial denaturation at 95°C for 3 min; 30 cycles at 95°C for 15 s, 62°C for 15 s, and 72°C for 15 s; and a final elongation at 72°C for 5 min. T7E1 analysis was performed as described for the *HTT* gene.

RNA Isolation and Reverse Transcription-Polymerase Chain Reaction

Total RNA was isolated from fibroblast cells using the TRI Reagent (BioShop; Burlington, Canada) according to the manufacturer's instructions. The RNA concentration was measured using spectrophotometer (DeNovix). A total of 700 ng of RNA was reverse transcribed at 55°C using Superscript III (Life Technologies) and random hexamer primers (Promega; Madison, WI, USA). The quality of the reverse transcription (RT) reaction was assessed through polymerase chain reaction (PCR) amplification of the GAPDH gene. Complementary DNA (cDNA) was used for quantitative polymerase chain reaction (qPCR) using SsoAdvanced™ Universal SYBR® Green Supermix (BIO-RAD, Hercules, CA, USA) with denaturation at 95°C for 30 s followed by 40 cycles of denaturation at 95°C for 15 s and annealing at 60°C for 30 s. The melt curve protocol was subsequently performed for 5 s at 65°C followed by 5 s increments at 0.5°C from 65°C to 95°C with *HTT*- or *GAPDH*-specific primers (sequences are listed in **Table S1**) on the CFX Connect™ Real-Time PCR Detection System (BIO-RAD). In order to avoid generation of two PCR products with different number of CAG repeats (two alleles of HD), primers used in qRT-PCR (HD_F and HD_R) were design to cover the HTT region downstream the CAG repeat tract. Data preprocessing and normalization were performed using BIO-RAD CFX Manager software (BIO-RAD). To confirm that the HTT transcript from the Cas9n-treated fibroblasts did not contain CAG repeats, cDNA was amplified with cDNAF and cDNAR primers flanking the repeat tract.

Western Blot Analysis

A total of 30 μg of protein was resolved on a Tris-acetate SDS-polyacrylamide gel (3–8%, NuPAGE™, Invitrogen, Carlsbad, CA, USA) in Tris-Acetate SDS Running buffer (Novex, Carlsbad, CA, USA) at 170 V at 4°C. After electrophoresis, the proteins were wet-transferred overnight to a nitrocellulose membrane (Sigma–Aldrich). The primary antibodies, namely, anti-huntingtin (1:1000, MAB2166, Millipore, Burlington, MA, USA) and anti-plectin (1:1000, ab83497, Abcam, Cambridge, UK), and the secondary antibodies, namely, the anti-mouse HRP conjugate (1:2000, A9917, Sigma–Aldrich) and anti-rabbit HRP conjugate (1:2000, 711-035-152, Jackson ImmunoResearch, West Grove, PA, USA) were used in a TBS/0.1% Tween-20 buffer containing 2.5% non-fat milk. The immunoreaction was detected using Western Bright Quantum HRP Substrate (Advansta, Menlo Park, CA, USA). The protein bands were scanned directly from the membrane using a camera and quantified using Gel-Pro Analyzer (Media Cybernetics).

Sanger Sequencing

DNA obtained from cell cultures transfected with plasmids was sequenced using a forward primer (HD1F). PCR products from DNA treated with Cas9 nickase pairs were separated in 1% agarose gel. Bands were extracted using the GeneJET Gel Extraction Kit (Thermo Scientific) and sequenced with the same HD1F primer.

Statistical Analysis

Statistical analysis was performed using GraphPad Prism v. 5.0 software. Data were analyzed using one-way ANOVA followed by Bonferroni's *post hoc* test (***$P < 0.0001$) with an arbitrary value of 1 assigned to the cells treated with the empty control plasmid.

RESULTS

Pre-screening of HTT_sgRNA Activity in HEK293T Cells

We designed 3 sgRNAs using *S. pyogenes* PAM sequences (NGG) located within the sequences flanking the CAG repeat tract in the *HTT* gene (HTT_sgRNA1, HTT_sgRNA3 and HTT_sgRNA4) and one sgRNA (HTT_sgRNA2) directly targeting the CAG repeats (**Figure 1A**). HTT_sgRNA2 was designed to use a non-canonical NAG PAM sequence that is known to be recognized by *S. pyogenes* Cas9 (Hsu et al., 2013; Zhang et al., 2014; Leenay et al., 2016). The first screening of the sgRNA activities was performed in easy-to-transfect HEK293T cells. The cells were transfected with plasmids expressing both the wt Cas9 protein and HTT_sgRNA (**Figure 1B**). The transfection efficiency, expressed as GFP-positive cells, was ∼80% (data not shown), and genomic

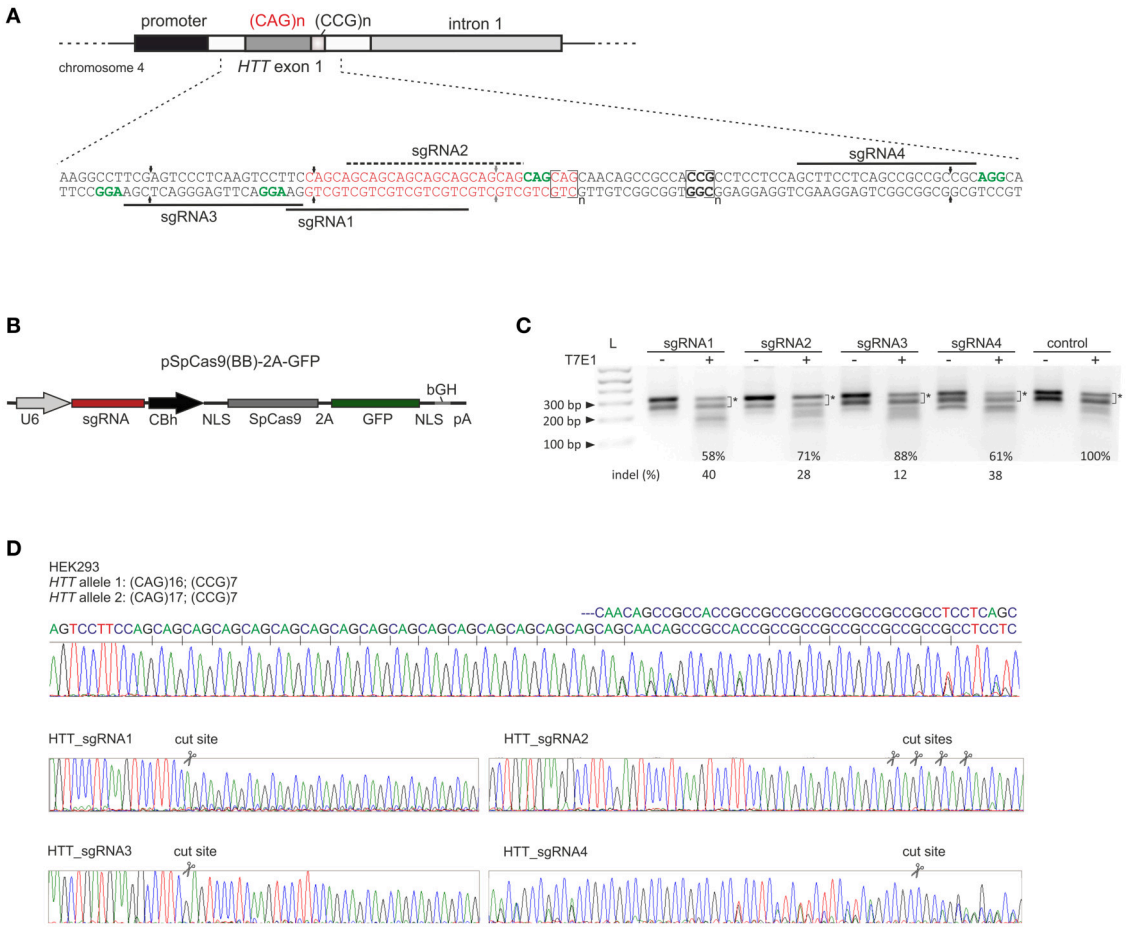


FIGURE 1 | Pre-screening of HTT_sgRNAs activity in HEK293T cells. **(A)** Polymorphic CAG repeats and the following CCG repeats are located in exon 1 of the *HTT* gene. sgRNA1, sgRNA3 and sgRNA4 were designed to target the 5′- and 3′-repeat flanking sequences, respectively. Appropriate PAM sequences are highlighted in green. SgRNA2 uses a non-canonical NAG PAM sequence and targets CAG repeats directly. **(B)** Illustration depicting the CRISPR/Cas9 expression plasmid used to co-express the SpCas9 protein together with the GFP reporter marker and sgRNA under the U6 promoter. NLS, nuclear localization signal. **(C)** Analysis of HTT_sgRNA and wt Cas9 activity in HEK293T cells by T7E1 mismatch assays. The signal intensities of the two main bands (marked with an asterisk) was measured. The loss of this signal in relation to the control (100%) was used to calculate the indel frequency (%) in the samples treated with T7E1 (+). The length of the main PCR product is ∼ 305 bp. Additional, faster migrating bands in samples non treated with T7E1 enzyme are secondary structure forms of the main product and their contribution is significantly reduced after denaturation of a sample directly before gel electrophoresis (see **Figure S1**). **(D)** Sanger sequencing of the *HTT* gene fragment included the CAG repeat tract. HEK293T cells are heterozygous in this locus with two alleles containing 16 and 17 CAG repeats. After *HTT* gene editing with CRISPR/Cas9, the sequence trace after the break site comprised a mixture of signals derived from the unmodified and modified DNA.

DNA was isolated 48 h post-transfection from unsorted cells. PCR amplification of the *HTT* gene region, including the CAG repeats, and subsequent T7E1 analysis resulted in the generation of multiple bands in both the treated and untreated control cells (**Figure 1C**). The HEK293T cells contain 16 and 17 CAG repeats in the two alleles of the *HTT* gene (**Figure 1D**). In addition, PCR products containing long stretches of repeated sequences may form various secondary structures (e.g., hairpins) in non-denaturing conditions (agarose gel electrophoresis) (**Figure S1**). Therefore, a determination of the exact indels frequency in this polymorphic, highly repetitive gene region was difficult using methods based on heteroduplex recognition by nucleases. However, Sanger sequencing (**Figure 1D**) and T7E1 analysis of the PCR products from a pool of transfected and non-transfected cells revealed that HTT_sgRNA1, HTT_sgRNA4 and HTT_sgRNA2 efficiently edited the *HTT* gene (∼40, 38, and ∼28% of indels, respectively) whereas HTT_sgRNA3 was the least active (∼12%) (**Figure 1C**). This result was consistent with the fact that the sequence composition of sgRNA (GC:AT content) may influence the editing efficiency (Moreno-Mateos MA 2015). The GC content for HTT_sgRNAs ranges from 57% for the least active HTT_sgRNA3 to 81% for HTT_sgRNA4.

Next, we analyzed the activities of the HTT_sgRNA pairs with the Cas9 nickase protein. HEK293T cells were transfected with a pair of plasmids encoding HTT_sgRNA1 and HTT_sgRNA2, HTT_sgRNA1 and HTT_sgRNA4, and HTT_sgRNA3 and HTT_sgRNA4. T7E1 analysis (**Figure 2A**) and Sanger sequencing (**Figure 2B**) confirmed that except for HTT_sgRNA1+2, the sgRNAs functioned in pairs and generated

shorter bands corresponding to HTT amplicons with 107-bp and 125-bp deletions for HTT_sgRNA1+4 and HTT_sgRNA3+4 pairs, respectively. As a result of the CAG repeat excision and frameshift mutation premature-translation-termination codon, PTC (TGA) was generated in the 3' region of *HTT* exon 1 (e.g., at the 44th codon of the 50-codon exon 1 for Cas9n_sgRNA1+4-treated cells) (**Figure S2**).

Analysis of Paired HTT_sgRNA/Cas9n Activity in Patient-Derived Fibroblasts

The most active pair of HTT_sgRNAs (sgRNA1 and sgRNA4) was electroporated into HD fibroblasts containing varied lengths of the CAG repeat tract: 21/44 CAGs, 17/68 CAGs and 21/151 CAGs. GFP-positive cells were sorted by FACS before DNA, RNA and protein isolation. We confirmed by PCR and Sanger sequencing of the PCR products that Cas9 nickase with the HTT_sgRNA1+4 pair efficiently excised the targeted region of *HTT* exon 1 in the patient-derived cell lines (**Figure S3**). The lengths of the excised DNA fragments were between 119 and 188 bp for both alleles of the GM04208 cell line, 107 and 260 bp for GM04281 and 119 and 509 for the GM09197 cell line. As expected, the HTT transcript also did not contain the targeted sequence, which was specifically excised from the DNA by the HTT_sgRNA1+4 Cas9 nickases (**Figure 3A**). A shorter, 545-bp PCR product was present in the three patients-derived cell lines; however, the editing efficiency was different, with the highest observed for the GM04281 cell line. Interestingly, the level of the HTT transcript did not change in the cells treated with paired nickases, suggesting that the transcript may have escaped the nonsense-mediated mRNA decay pathway **(Figure 3B)**. Notably, the newly generated stop codon (UGA) was localized ∼20 bp from the exon/exon junction and may have been occupied by the exon-exon junction complex (EJC) (**Figure S2**). Despite the presence of a shortened HTT transcript, the HTT protein level was efficiently reduced by 82% in the GM04281 cell line, 68% in the GM04208 cell line, and 71% in the GM09197 cell line (**Figures 3C,D**). This data accurately reflects the results of the RT-PCR analysis (**Figure 3A**) and indicates that the length of the CAG repeat tract does not influence the excision efficiency of Cas9n. The prematurely terminated translation product (43 amino-acid protein) was not detected with the use of the N-terminal huntingtin antibody by western blot (data not shown).

Assessment of Off-Target Effects

In silico analysis using the CRISPOR tool (Haeussler et al., 2016) predicted 13 exonic off-target sites for HTT_sgRNA1 (with score > 10.00), 18 sites for HTT_sgRNA3 and 196 for HTT_sgRNA4. Notably, more than 98% of the exonic HTT_sgRNA4 off-targets had 3 to 4 mismatches with the target sequence (**Table S2**). Specificity score that measures the uniqueness of a guide in the genome is low for HTT_sgRNAs, because target sequence is composed of repetitive sequences (6 for HTT_sgRNA1, 66 for HTT_sgRNA3 and 40 for HTT_sgRNA4). HTT_sgRNA2 with a non-canonical NAG PAM comprised a repeated sequence and theoretically targeted every CAG repeat tract longer than

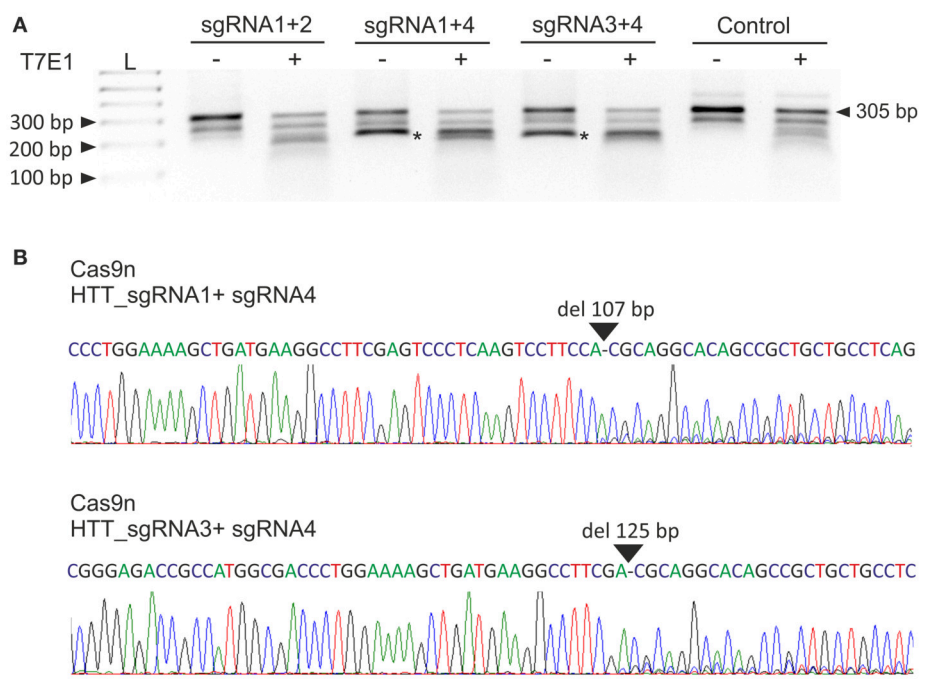


FIGURE 2 | Excision of the CAG repeats from the HTT gene with paired Cas9 nickases. **(A)** DNA from HEK293T cells treated with Cas9 nickase and the HTT_sgRNA pairs was PCR amplified and subjected to T7E1 analysis. The length of the unmodified PCR fragment is 305 bp, whereas the edited products are shortened to < 200 bp (marked with an asterisk). **(B)** A representative Sanger sequencing electropherogram showing the deletion of 107 and 125 bp in HTT_sgRNA1+4- and HTT_sgRNA1+3-treated cells, respectively (marked with an arrow).

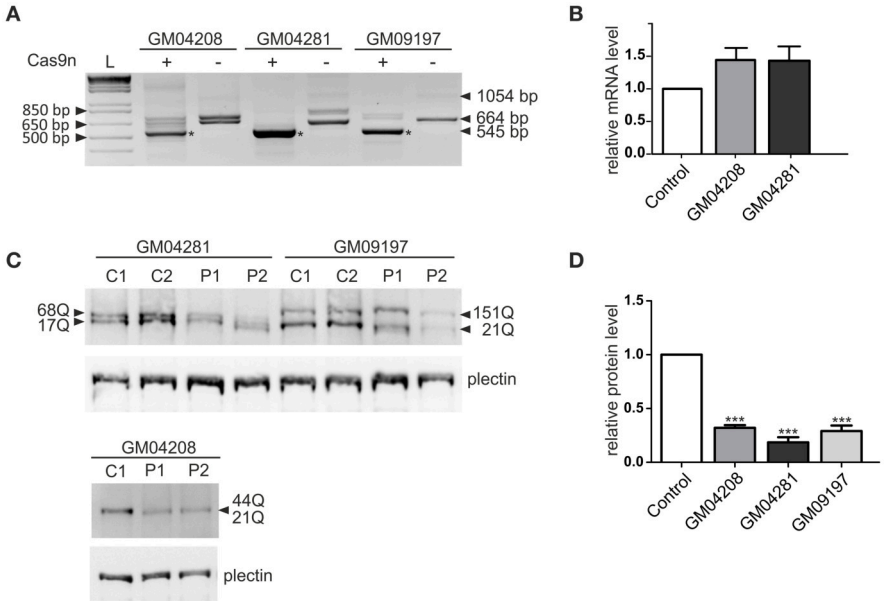


FIGURE 3 | Huntingtin inactivation via the Cas9 nickase pair in patient-derived fibroblasts. **(A)** RT-PCR analysis of the Cas9n/HTT_sgRNA1+4-edited (+) and non-edited (–) *HTT* gene product in the three human HD cell lines containing 21/44 (GM04208), 17/68 (GM04281) and 21/151 (GM09197) CAG repeats. As a result of repeat excision, shorter PCR products (∼545 bp, marked with an asterisk) are present. **(B)** RT-qPCR analysis of the HTT mRNA levels in the human fibroblast cell lines transfected with the Cas9n/HTT_sgRNA1 and sgRNA4 plasmid pairs. All samples are normalized to human GAPDH, and the results are the mean (± SEM) relative to the cells transfected with the control Cas9n plasmid (one-way ANOVA followed by Bonferroni's *post hoc* test; the difference was non-significant). **(C)** Representative western blots showing the HTT protein levels in the control fibroblasts treated with the empty Cas9n plasmids, expressing only Cas9 protein and not sgRNA (C1 and C2) and in the Cas9n/HTT_sgRNA1+4-treated cells (P1 and P2). Plectin was used as the loading control. The lengths of the polyQ tracts in both alleles of the HTT protein in each cell line are marked with an arrow. **(D)** The quantification of the huntingtin levels in the three human HD cell lines transfected with the Cas9n/HTT_sgRNA1+4 plasmid pair relative to the plectin levels. The results indicate the mean (± SEM) relative to the cells transfected with the control Cas9n plasmid ($n = 3$; one way ANOVA followed by Bonferroni's *post hoc* test; $^{***}p < 0.0001$).

7 units. Notably, these predictions generally applied to the wt CRISPR/Cas9 activity since Cas9 nickases cut only one DNA strand that is faithfully repaired by HDR. In addition, the maximal cleavage efficiency of paired Cas9 nickases has previously been observed at sites with the tail-to-tail orientation separated by 10–30 bp (Shen et al., 2014). Potential off-target activity for Cas9n/HTT_sgRNA1+4 pair was expected to be rare as similar sequences were unlikely to occur close together elsewhere in the genome. Nonetheless, the 4 selected off-target regions for each HTT_sgRNA were PCR-amplified and analyzed with T7E1 assays (**Figure S4**). The *TEX13A* and *ZFHX3* genes and *TJP2* and *FBXW7* genes were tested for HTT_sgRNA1 and HTT_sgRNA4, respectively. We used DNA from HEK293T cells treated with plasmids expressing the HTT_sgRNA1+4 Cas9 nickase pair and wt Cas9/sgRNAs. Non-specific activity of the CRISPR/Cas9 system was not detected in any of the tested off-target sites.

DISCUSSION

To date, multiple therapeutic approaches have been described for the treatment of HD and other polyQ diseases' however, these approaches suffer from specific limitations that hinder their introduction to the clinic (reviewed in Keiser et al., 2016; Esteves et al., 2017; Wild and Tabrizi, 2017). In addition, the role of huntingtin in cell physiology and pathology is not fully understood (Saudou and Humbert, 2016), and therefore, strategies using selective silencing of the mutant allele alone and non-allele-selective silencing of both alleles are being developed in parallel. It has been shown using RNAi and antisense oligonucleotides that the knockdown of huntingtin, either the mutant or both mutant and normal is beneficial in mouse models of HD (Harper et al., 2005; Boudreau et al., 2009; Kordasiewicz et al., 2012). Recently, the CRISPR/Cas9 system was used to permanently inactivate the *HTT* gene, by using a pair of sgRNAs flanking the CAG/CTG repeats in a transgenic mouse model of HD (HD140Q knock-in) (Yang et al., 2017). Stereotactic injection of AAVs expressing sgRNAs and SpCas9 into the striata of adult mice resulted in the depletion of huntingtin aggregates in the brain, thereby alleviating motor deficits and neuropathological symptoms.

In our study, we present another repeat-depletion strategy to inactivate the *HTT* gene in which we further improve the approach by using a nickase version of Cas9 that is known to be more specific and safe than the wt Cas9. The efficiency of paired Cas9 nickase editing depends on the activity of two sgRNAs and the length of the target sequence between the two sgRNAs (Mali et al., 2013; Ran et al., 2013). We demonstrate that the pair of HTT_sgRNA/Cas9n is able to efficiently and specifically excise the repeat-containing fragment of exon 1 in three HD

patient-derived cell lines differing in CAG repeat length. We show that the CAG repeat length did not influence the cutting efficiency and specificity, and the HTT protein level is reduced by ∼70% in all tested models. Notably, in the case of the GM09197 cell line containing 151 CAG repeats in the mutant allele, the HTT_sgRNAs were separated by nearly 500 bp. We confirmed the specificity and safety of the paired nickase strategy by testing selected off-target loci with T7E1 mismatch detection assays.

The mechanism of this precise repeat excision and DNA repair (without scars), atypical for NHEJ is poorly known and needs further studies. However, it has been reported previously that DSBs generated by CRISPR/Cas9 near a long stretch of CTG/CAG repeats in myotonic dystrophy type 1 (DM1) locus can induce deletion of the entire repeat region (Van Agtmaal et al., 2017). Even single DSB in the region flanking the repeated sequence was sufficient to generate clean loss of repeats. Contraction of CAG/CTG repeats was also observed for ZFN and TALEN—treated human and yeast cells (Mittelman et al., 2009; Richard et al., 2014).

Interestingly, in our study the level of the shortened HTT transcript did not change, suggesting that the transcript may be NMD-resistant. The *HTT* gene contains 67 exons and has three isoforms of mRNA transcripts (Romo et al., 2017); the two predominant forms are 10,366 and of 13,711 bp (Lin et al., 1993). The longer transcript differs by an additional 3′ UTR sequence of 3,360 bp that affects mRNA localization, stability, and translation (Di Giammartino et al., 2011). A previous report showed that targeting the HTT exon 1-intron junction with CRISPR/Cas9 reduced the mRNA level by ∼50% in BM-MSCs derived from the YAC128 mouse model (Kolli et al., 2017). In another study, a large deletion of approximately 44 kb of DNA using wt Cas9 and a pair of sgRNAs targeting the upstream promoter region and intron 3 resulted in the complete abrogation of mRNA and HTT protein synthesis (Shin et al., 2016). HTT mRNA resistance to NMD, observed in our study, may result from the specific localization of CRISPR/Cas9n-generated PTC in exon 1 of the multi-exonic *HTT* gene. In addition, a UGA stop codon is localized to position occupied by the EJC (∼20–24 nt upstream of the exon/exon junction), which serves to orient the NMD machinery and may be masked during the "pioneer round" of translation (Popp and Maquat, 2016). A previous report showed that β-globin transcript containing PTCs in exon 1 of three-exonic gene is NMD–resistant; however, the influence of the nonsense codon localization within transcripts needs to be clarified (Thermann et al., 1998; Inácios et al., 2004; Peixeiro et al., 2011).

Genome editing with the use of a more universal CAG repeat-targeting strategy is still challenging due to the lack of specific PAM recognized by targeted nucleases, off-targeting induced by sgRNA comprising repeats and problems with the selective inactivation of mutant alleles alone. Similar problems have already been overcome by antisense and RNAi technologies (Hu et al., 2009; Yu et al., 2012; Fiszer et al., 2013). In our study CAG repeat targeting with Cas9n and HTT_sgRNA pair composed of sgRNA2 (non-canonical NAG PAM) and sgRNA1 was ineffective. However, the in-frame shortening of the CAG repeat tract with the use of genome editing tools would be the most desired and universal approach and is our goal for future studies.

AUTHOR CONTRIBUTIONS

MO and MD contributed to the study design. MD, MO, and WJ performed experiments. MO, MD, and WK contributed to the data analysis, writing, and editing of the manuscript.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Science Center (2015/18/E/NZ2/00678) and from the quality-promoting subsidy under the Leading National Research Center (KNOW) program for 2014–2018.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2018.00075/full#supplementary-material

REFERENCES

An, M. C., O'Brien, R. N., Zhang, N., Patra, B. N., De La Cruz, M., Ray, A., et al. (2014). Polyglutamine disease modeling: epitope based screen for homologous recombination using CRISPR/Cas9 System. *PLoS Curr*. 6:ecurrents.hd.0242d2e7ad72225efa72f6964589369a. doi: 10.1371/currents.hd.0242d2e7ad72225efa72f6964589369a

An, M. C., Zhang, N., Scott, G., Montoro, D., Wittkop, T., Mooney, S., et al. (2012). Genetic correction of huntington's disease phenotypes in induced pluripotent stem cells. *Cell Stem Cell* 11, 253–263. doi: 10.1016/j.stem.2012.04.026

Bates, G. P., Dorsey, R., Gusella, J. F., Hayden, M. R., Kay, C., Leavitt, B. R., et al. (2015). Huntington disease. *Nat. Rev. Dis. Prim*. 1:15005. doi: 10.1038/nrdp.2015.5

Boudreau, R. L., McBride, J. L., Martins, I., Shen, S., Xing, Y., Carter, B. J., et al. (2009). Nonallele-specific silencing of mutant and wild-type Huntingtin demonstrates therapeutic efficacy in Huntington's disease mice. *Mol. Ther*. 17, 1053–1063. doi: 10.1038/mt.2009.17

Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., et al. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res*. 24, 132–141. doi: 10.1101/gr.162339.113

Di Giammartino, D. C., Nishida, K., Manley, J. L. L., Di Giammartino, D. C., Nishida, K., and Manley, J. L. L. (2011). Mechanisms and consequences of alternative polyadenylation. *Mol. Cell* 43, 853–866. doi: 10.1016/j.molcel.2011.08.017

Duyao, M., Ambrose, C., Myers, R., Novelletto, A., Persichetti, F., Frontali, M., et al. (1993). Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nat. Genet*. 4, 387–392. doi: 10.1038/ng0893-387

Esteves, S., Duarte-Silva, S., and Maciel, P. (2017). Discovery of therapeutic approaches for polyglutamine diseases: a summary of recent efforts. *Med. Res. Rev*. 37, 860–906. doi: 10.1002/med.21425

Fiszer, A., Olejniczak, M., Galka-Marciniak, P., Mykowska, A., and Krzyzosiak, W. J. (2013). Self-duplexing CUG repeats selectively inhibit mutant huntingtin expression. *Nucleic Acids Res*. 41, 10426–10437. doi: 10.1093/nar/gkt825

Fiszer, A., Olejniczak, M., Switonski, P. M., Wroblewska, J. P., Wisniewska-Kruk, J., Mykowska, A., et al. (2012). An evaluation of oligonucleotide-based therapeutic strategies for polyQ diseases. *BMC Mol. Biol*. 13:6. doi: 10.1186/1471-2199-13-6

Garriga-Canut, M., Agustín-Pavón, C., Herrmann, F., Sánchez, A., Dierssen, M., Fillat, C., et al. (2012). Synthetic zinc finger repressors reduce mutant

huntingtin expression in the brain of R6/2 mice. *Proc. Natl. Acad. Sci. U.S.A.* 109, E3136–E3145. doi: 10.1073/pnas.1206506109

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J. B., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* 17:148. doi: 10.1186/s13059-016-1012-2

Harper, S. Q., Staber, P. D., He, X., Eliason, S. L., Martins, I. H., Mao, Q., et al. (2005). RNA interference improves motor and neuropathological abnormalities in a Huntington's disease mouse model. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5820–5825. doi: 10.1073/pnas.0501507102

Heman-Ackah, S. M., Bassett, A. R., and Wood, M. J. A. (2016). Precision modulation of neurodegenerative disease-related gene expression in human iPSC-derived neurons. *Sci. Rep.* 6:28420. doi: 10.1038/srep28420

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31, 827–832. doi: 10.1038/nbt.2647

Hu, J., Matsui, M., Gagnon, K. T., Schwartz, J. C., Gabillet, S., Arar, K., et al. (2009). Allele-specific silencing of mutant huntingtin and ataxin-3 genes by targeting expanded CAG repeats in mRNAs. *Nat. Biotechnol.* 27, 478–484. doi: 10.1038/nbt.1539

Inácios, Â., Silva, A. L., Pinto, J., Ji, X., Morgado, A., Almeida, F., et al. (2004). Nonsense mutations in close proximity to the initiation codon fail to trigger full nonsense-mediated mRNA decay. *J. Biol. Chem.* 279, 32170–32180. doi: 10.1074/jbc.M405024200

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829

Jordan, M., Schallhorn, A., and Wurm, F. M. (1996). Transfecting mammalian cells: optimization of critical parameters affecting calcium-phosphate precipitate formation. *Nucleic Acids Res.* 24, 596–601. doi: 10.1093/nar/24.4.596

Keiser, M. S., Kordasiewicz, H. B., and McBride, J. L. (2016). Gene suppression strategies for dominantly inherited neurodegenerative diseases: lessons from Huntington's disease and spinocerebellar ataxia. *Hum. Mol. Genet.* 25, R53–R64. doi: 10.1093/hmg/ddv442

Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., et al. (1994). Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). *Nat. Genet.* 6, 9–13. doi: 10.1038/ng0194-9

Kolli, N., Lu, M., Maiti, P., Rossignol, J., and Dunbar, G. L. (2017). CRISPR-Cas9 mediated gene-silencing of the mutant Huntingtin gene in an *in vitro* model of Huntington's disease. *Int. J. Mol. Sci.* 18:754. doi: 10.3390/ijms18040754

Kordasiewicz, H. B., Stanek, L. M., Wancewicz, E. V., Mazur, C., McAlonis, M. M., Pytel, K. A., et al. (2012). Sustained therapeutic reversal of Huntington's disease by transient repression of Huntingtin synthesis. *Neuron* 74, 1031–1044. doi: 10.1016/j.neuron.2012.05.009

Leenay, R. T., Maksimchuk, K. R., Slotkowski, R. A., Agrawal, R. N., Gomaa, A. A., Briner, A. E., et al. (2016). Identifying and visualizing functional PAM diversity across CRISPR-cas systems. *Mol. Cell* 62, 137–147. doi: 10.1016/j.molcel.2016.02.031

Lin, B., Rommens, J. M., Graham, R. K., Kalchman, M., Macdonald, H., Nasir, J., et al. (1993). Differential 3' polyadenylation of the huntington disease gene results in two mRNA species with variable tissue expression. *Hum. Mol. Genet.* 2, 1541–1545. doi: 10.1093/hmg/2.10.1541

Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., et al. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* 31, 833–838. doi: 10.1038/nbt.2675

Merienne, N., Vachey, G., de Longprez, L., Meunier, C., Zimmer, V., Perriard, G., et al. (2017). The self-inactivating KamiCas9 system for the editing of CNS disease genes. *Cell Rep.* 20, 2980–2991. doi: 10.1016/j.celrep.2017.08.075

Mittelman, D., Moye, C., Morton, J., Sykoudis, K., Lin, Y., Carroll, D., et al. (2009). Zinc-finger directed double-strand breaks within CAG repeat tracts promote repeat instability in human cells. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9607–9612. doi: 10.1073/pnas.0902420106

Monteys, A. M., Ebanks, S. A., Keiser, M. S., and Davidson, B. L. (2017). CRISPR/Cas9 editing of the mutant huntingtin allele *in vitro* and *in vivo*. *Mol. Ther.* 25, 12–23. doi: 10.1016/j.ymthe.2016.11.010

Paulson, H. L., Shakkottai, V. G., Clark, H. B., and Orr, H. T. (2017). Polyglutamine spinocerebellar ataxias — from genes to potential treatments. *Nat. Rev. Neurosci.* 18, 613–626. doi: 10.1038/nrn.2017.92

Peixeiro, I., Silva, A. L., and Romão, L. (2011). Control of human β-globin mRNA stability and its impact on beta-thalassemia phenotype. *Haematologica* 96, 905–913. doi: 10.3324/haematol.2010.039206

Popp, M. W., and Maquat, L. E. (2016). Leveraging rules of nonsense-mediated mRNA decay for genome engineering and personalized medicine. *Cell* 165, 1319–1332. doi: 10.1016/j.cell.2016.05.053

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., et al. (2013). Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell* 154, 1380–1389. doi: 10.1016/j.cell.2013.08.021

Richard, G. F., Viterbo, D., Khanna, V., Mosbach, V., Castelain, L., and Dujon, B. (2014). Highly specific contractions of a single CAG/CTG trinucleotide repeat by TALEN in yeast. *PLoS ONE* 9:e95611. doi: 10.1371/journal.pone.0095611

Romo, L., Ashar-Patel, A., Pfister, E., and Aronin, N. (2017). Alterations in mRNA 3′ UTR isoform abundance accompany gene expression changes in human Huntington's disease brains. *Cell Rep.* 20, 3057–3070. doi: 10.1016/j.celrep.2017.09.009

Saudou, F., and Humbert, S. (2016). The biology of Huntingtin. *Neuron* 89, 910–926. doi: 10.1016/j.neuron.2016.02.003

Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., et al. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* 11, 399–402. doi: 10.1038/nmeth.2857

Shin, J. W., Kim, K.-H., Chao, M. J., Atwal, R. S., Gillis, T., MacDonald, M. E., et al. (2016). Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9. *Hum. Mol. Genet.* 25, 4566–4576. doi: 10.1093/hmg/ddw286

Spada, A. R., La Wilson, E. M., Lubahn, D. B., Harding, A. E., and Fischbeck, K. H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 352, 77–79. doi: 10.1038/352077a0

Thermann, R., Neu-Yilik, G., Deters, A., Frede, U., Wehr, K., Hagemeier, C., et al. (1998). Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J.* 17, 3484–3494. doi: 10.1093/emboj/17.12.3484

Trevino, A. E., and Zhang, F. (2014). Genome editing using cas9 nickases. *Methods Enzymol.* 546, 161–174. doi: 10.1016/B978-0-12-801185-0.00008-8

Van Agtmaal, E. L., André, L. M., Willemse, M., Cumming, S. A., Van Kessel, I. D. G., Van Den Broek, W. J. A. A., et al. (2017). CRISPR/Cas9-Induced (CTG,CAG) n repeat instability in the myotonic dystrophy type 1 locus: implications for therapeutic genome editing. *Mol. Ther.* 25, 24–43. doi: 10.1016/j.ymthe.2016.10.014

Wild, E. J., and Tabrizi, S. J. (2017). Therapies targeting DNA and RNA in Huntington's disease. *Lancet Neurol.* 16, 837–847. doi: 10.1016/S1474-4422(17)30280-6

Yang, S., Chang, R., Yang, H., Zhao, T., Hong, Y., Kong, H. E., et al. (2017). CRISPR/Cas9-mediated gene editing ameliorates neurotoxicity in mouse model of Huntington's disease. *J. Clin. Invest.* 127, 2719–2724. doi: 10.1172/JCI92087

Yu, D., Pendergraff, H., Liu, J., Kordasiewicz, H. B., Cleveland, D. W., Swayze, E. E., et al. (2012). Single-stranded RNAs use RNAi to potently and allele-selectively inhibit mutant huntingtin expression. *Cell* 150, 895–908. doi: 10.1016/j.cell.2012.08.002

Zhang, Y., Ge, X., Yang, F., Zhang, L., Zheng, J., Tan, X., et al. (2014). Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Sci. Rep.* 4:5405. doi: 10.1038/srep05405

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Dabrowska, Juzwa, Krzyzosiak and Olejniczak. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Supplementary Material

"Precise excision of the CAG tract from the Huntingtin gene by Cas9 nickases"
Dabrowska M. et al., Frontiers in Neuroscience

Supplementary Table S1. Oligonucleotides used in the study

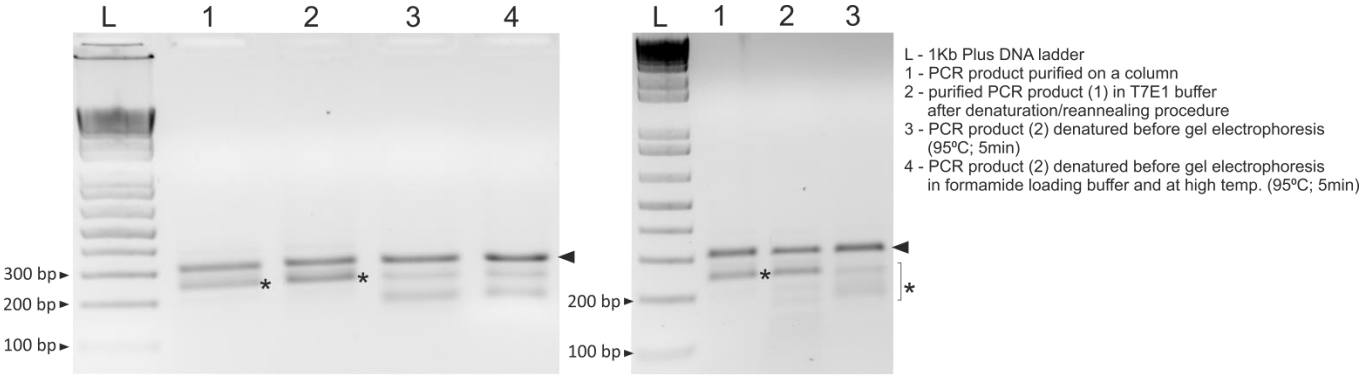
Oligonucleotide ID	**Sequence (5'-3')**	**Description**
sgRNA1s	CACCGCTGCTGCTGCTGCTGCTGGA	oligo for HTT_sgRNA1 plasmid construction
sgRNA1a	AAACTCCAGCAGCAGCAGCAGCAGC	oligo for HTT_sgRNA1 plasmid construction
sgRNA2s	CACCGAGCAGCAGCAGCAGCAGCAG	oligo for HTT_sgRNA2 plasmid construction
sgRNA2a	AAACCTGCTGCTGCTGCTGCTGCTC	oligo for HTT_sgRNA2 plasmid construction
sgRNA3s	CACCGGAAGGACTTGAGGGACTCGA	oligo for HTT_sgRNA3 plasmid construction
sgRNA3a	AAACTCGAGTCCCTCAAGTCCTTCC	oligo for HTT_sgRNA3 plasmid construction
sgRNA4s	CACCGGCTTCCTCAGCCGCCGCCGC	oligo for HTT_sgRNA4 plasmid construction
sgRNA4a	AAACGCGGCGGCGGCTGAGGAAGCC	oligo for HTT_sgRNA4 plasmid construction
U6-Fwd	GAGGGCCTATTTCCCATGATTCC	Sequencing primer
HD1F	CCGCTCAGGTTCTGCTTTTA	PCR primer, sequencing primer
HD1R	GGCTGAGGCAGCAGCGGCTG	PCR primer
GAPDH_F	GAAGGTGAAGGTCGGAGTC	qRT-PCR primer
GAPDH_R	GAAGATGGTGATGGGATTTC	qRT-PCR primer
HD_F	CGACAGCGAGTCAGTGATTG	qRT-PCR primer
HD_R	ACCACTCTGGCTTCACAAGG	qRT-PCR primer
cDNAF	CCCTGGAAAAGCTGATGAAG	primer for *HTT* cDNA, sequencing primer
cDNAR	TCTTCGGGTCTCTTGCTTGT	primer for *HTT* cDNA
ZFHX3-F	CCAAATAAACCGTCCTCAGC	primer for ZFHX gene- sgRNA1 off-target
ZFHX3-R	TTCCCTTTGTGTGCCTTTTC	primer for ZFHX gene- sgRNA1 off-target
TEX13A-F	CGTCCTACCCTGCTTAGTGC	primer for TEX13A gene-sgRNA1 off-target
TEX13A-R	GGTTCGTGGTTCCAGAGAAA	primer for TEX13A gene- sgRNA1 off-target
TJP2-F	GTAGCGGCCAATTTGACAGT	primer for TJP2 gene- sgRNA4 off-target
TJP2-R	CACAAGGAGGCACTTACGC	primer for TJP2 gene- sgRNA4 off-target
FBXW7-F	CACAGAGCGAGGGAGACAG	primer for FBXW7 gene- sgRNA4 off-target
FBXW7-R	CCTCCTCAGCGTTCTCTCAC	primer for FBXW7 gene- sgRNA4 off-target

Supplementary Table S2. Predicted exonic off-targets regions for HTT_sgRNA1 and HTT_sgRNA4 (http://crispor.tefor.net)

Off-target	Guide sequence **sgRNA_1** CTGCTGCTGCTGCTGCTGGA	**PAM** AGG	Chromosome	Gene	Strand	Mismatches
1	CTGCTGCTGCTGCTGCTGG**G**	GGG	chr16	*ZFHX3*	+	1
2	CTGCTGCTGCTGCTGCTGG**G**	GGG	chr19/3'UTR	*DMPK*	-	1
3	CTGCTGCTGCTGCTGCTG**C**A	AGG	chr15	*SEMA6D*	+	1
4	CTGCTGCTGCTGCTGCTGG**C**	GGG	chr2	*APOB/ex1*	-	1
5	CTGCTGCTGCTGCTGCTGG**C**	GGG	chr1	*SDC3/ex1*	-	1
6	CTGCTGCTGCTGCTGCTGG**C**	CGG	chr4	*SDAD1*	-	1
7	CTGCTGCTGCTGCTGCTGG**C**	AGG	chr11	*AP2A2*	+	1
8	**T**TGCTGCTGCTGCTGCTGG**C**	TGG	chr22	*TCF20*	+	2
9	CTGCTGCTGCTGCTGCTGGA	GGA	chr1	*NOS1AP*	-	0
10	CTGCTG**A**TGCTGCTGCTGGA	TGA	chr2	*SOX11*	-	1
11	CTGCTG**A**TGCTGCTGCTGGA	TGA	chr2	*HDAC4*	+	1
12	CTGCTGCTG**G**TGCTGCTGGA	GGA	chrX	*TEX13A*	-	1
13	CTGCTGCTG**G**TGCTGCTGGA	GGA	chrX	*TEX13B*	-	1

Off-target	Guide sequence **sgRNA_4** GCTTCCTCAGCCGCCGCCGC	**PAM** AGG	Chromosome	Gene	Strand	Mismatches
1	**T**CTTCCTCA**T**CC**A**CCGCC**A**C	TGG	chr12	*PXN-AS1*	-	4
2	GCT**C**C**T**CCAGCCGCCGCCGC	TGG	chr8	*PLEC*	+	3
3	GCTTCC**GG**AGCCGCCGCCGC	AGG	chr16	*CDH5*	-	2
4	GCT**G**CC**G**CAGCCGCCGCCGC	AGG	chr12	*MBD6*	-	2
5	**AT**TTCCT**GG**GCCGCCGCCGC	CGG	chr14	*DIO3*	+	4
6	GC**C**TC**T**T**T**AGCC**A**CCGCCGC	CGG	chr12	*YBX3*	-	4
7	GCTTCC**C**CAGC**A**GCC**A**C**T**GC	TGG	chr17	*CDC42EP4*	+	4
8	GCT**G**CC**A**C**C**GCCGCCGCCGC	AGG	chr20	*DUSP15/TTLL9*	+	3
9	G**G**T**C**CCT**G**AGCCGCCGCCGC	GGG	chr17	*MMP28*	-	3
10	**T**C**C**TCCTCAGCCGCCGCC**T**C	AGG	chr12	CLEC1A	-	3
11	GCT**G**CC**G**C**C**GCCGCCGCCGC	TGA	chr4	FBXW7	-	3
12	GCT**GA**C**G**C**C**GCCGCCGCCGC	GGG	chr9	TJP2	+	4

FIGURE S1



Analysis of a non-specific band generated during agarose gel electrophoresis of HTT PCR product. Genomic DNA from HEK293 cells (controls from Cas9 experiments, see Fig. 1C) was amplified using Phusion High-Fidelity PCR Master Mix with primers HD1F and HD1R spanning CAG repeats in exon 1 of the HTT gene. The two-step PCR amplification program was used as follows: an initial denaturation at 98°C for 3 min; 12 cycles at 98°C for 15 s, 72°C for 15 s; 21 cycles at 98°C for 15 s, 62°C for 15 s, and 72°C for 15 s; and a final elongation at 72°C for 5 min. PCR products were purified using the GeneJET PCR Purification Kit. 400 ng of the purified PCR product (1), PCR product after T7E1 annealing reaction (2) or PCR product after denaturation at high temperature (3) and formamide (4) were separated in 1.3% agarose gels and detected using G-BOX. Two gels represent two independent experiments. The main product (~ 305 bp) is indicated with an arrowhead. Faster migrating bands are secondary structure forms of the main product (marked with a star) and their contribution is significantly reduced after denaturation of a sample directly before gel electrophoresis.

FIGURE S2

The sequence targeted by Cas9n/HTT_sgRNAs within exon 1 of the *HTT* gene

a) Human huntingtin 3144 aa, 345kDa

ATG GCG ACC CTG GAA AAG CTG ATG AAG GCC TTC GAG TCC CTC AAG
TCC TTC **CA**G CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG
CAG CAG CAG CAG CAG CAG CAA CAG CCG CCA CCG CCG CCG CCG CCG
CCG CCG CCT CCT CAG CTT CCT CAG CCG CCG C**CG** CAG GCA CAG CCG
CTG CTG CCT CAG CCG CAG CCG CCC CCG CCG CCG CCC CCG CCG CCA
CCC GGC CCG GCT GTG GC**T GA**G GAG CCG CTG CAC CGA CCAAAGAAAGA
ACTTTCAG…

Cas9n/HTT_sgRNA1+4; 43 aa 4,73kDa

ATG GCG ACC CTG GAA AAG CTG ATG AAG GCC TTC GAG TCC CTC AAG
TCC TTC **CAC G**CA GGC ACA GCC GCT GCT GCC TCA GCC GCA GCC GCC
CCC GCC GCC GCC CCC GCC GCC ACC CGG CCC GGC TGT GGC **TGA** GGA
GCC GCT GCA CCG ACCAAAGAAAGAACTTTCAG…

b) Human huntingtin 3144 aa, 345kDa

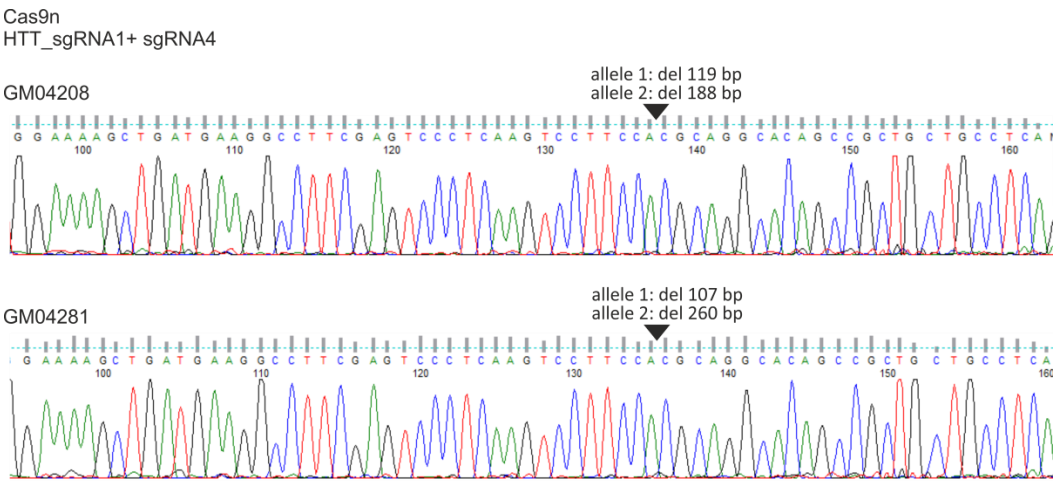
ATG GCG ACC CTG GAA AAG CTG ATG AAG GCC TTC **GA**G TCC CTC AAG
TCC TTC CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG
CAG CAG CAG CAG CAG CAG CAA CAG CCG CCA CCG CCG CCG CCG CCG
CCG CCG CCT CCT CAG CTT CCT CAG CCG CCG C**CG** CAG GCA CAG CCG
CTG CTG CCT CAG CCG CAG CCG CCC CCG CCG CCG CCC CCG CCG CCA
CCC GGC CCG GCT GTG GC**T GA**G GAG CCG CTG CAC CGA CCAAAGAAAGA
ACTTTCAG…

Cas9n/HTT_sgRNA3+4; 37 aa, 4kDa

ATG GCG ACC CTG GAA AAG CTG ATG AAG GCC TTC **GAC G**CA GGC ACA
GCC GCT GCT GCC TCA GCC GCA GCC GCC CCC GCC GCC GCC CCC GCC
GCC ACC CGG CCC GGC TGT GGC **TGA** GGA GCC GCT GCA CCG ACCAAAG
AAAGAACTTTCAG…

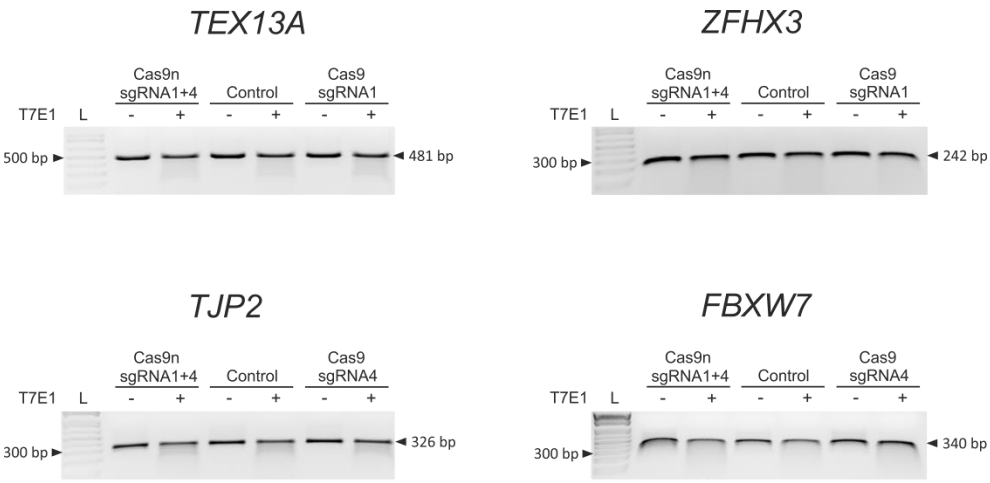
The sequence removed from exon 1 of the *HTT* gene by paired Cas9 nickases is highlighted in red and the nucleotides directly flanking the Cas9n-induced nicks are underlined and indicated in bold. As a result of the CAG repeat excision and frameshift mutation, a TGA STOP codon is generated (highlighted in yellow). The intronic sequence is indicated in blue.

FIGURE S3



Sanger sequencing analysis of *HTT* gene editing in human fibroblast cell lines.

FIGURE S4



Analysis of potential off-target loci by T7E1 mismatch detection assay. DNA from HEK293T cells treated with Cas9n/HTT_sg1+4, Cas9/HTT_sgRNA1, and Cas9/HTT_sgRNA4 and from the control cells transfected with empty plasmid without sgRNAs was amplified using primers specific for HTT_sgRNA1 (TEX13A, ZFHX3) and HTT_sgRNA4 (TJP2, FBXW7) off-target genes. Next, T7E1 analysis was performed to examine cleavage activity in the potential off-target sites. Purified PCR products for appropriate genes were treated (+) and untreated (-) with the T7E1 enzyme and separated on agarose gels with a 1 kb Plus DNA ladder (L).

Chapter 17

Gene Therapy for Huntington's Disease Using Targeted Endonucleases

Magdalena Dabrowska and Marta Olejniczak

Abstract

Huntington's disease (HD) is a hereditary neurological disorder caused by expansion of the CAG repeat tract in the huntingtin gene (*HTT*). The mutant protein with a long polyglutamine tract is toxic to cells, especially neurons, leading to their progressive degeneration. Similar to many other monogenic diseases, HD is a good target for gene therapy approaches, including the use of programmable endonucleases. Here, we describe a protocol for *HTT* gene knock out using a modified Cas9 protein (nickase, Cas9n) and a pair of sgRNAs flanking the repeats. Recently, we showed that excision of the CAG repeat tract resulted in a frameshift mutation and premature translation termination. As a model, we used HD patient-derived fibroblasts electroporated with a pair of plasmid vectors expressing CRISPR-Cas9n tools. Efficient *HTT* inactivation independent of the CAG tract length was confirmed by Western blotting. A modified version of this protocol involving precise excision of the CAG repeats and insertion of a new DNA sequence by homology directed repair may also be used for the generation of new isogenic cellular models of HD.

Key words Gene therapy, Polyglutamine diseases, CRISPR, Cas9, CAG repeats, HDR

1 Introduction

Huntington's disease (HD) is a hereditary neurological disorder that is caused by a mutation in the first exon of the *HTT* gene. HD patients contain abnormally expanded CAG repeats that translate into a polyglutamine (polyQ) tract in one allele of the *HTT* gene. The accumulation of polyQ-rich huntingtin protein induces a toxic gain-of-function property and affects various cellular activities. It is also postulated that long CAG repeats in transcripts may form a hairpin structure that is responsible for RNA toxicity. HD symptoms involve the loss of neurons in the striatum and cortex accompanied by reactive gliosis and astrocytosis, which lead to progressive movement abnormalities and dementia [1].

Despite many years of research searching for the best therapeutic strategy and considerable progress in the development of antisense and RNA interference (RNAi) technologies, HD is still

Guy-Franck Richard (ed.), *Trinucleotide Repeats: Methods and Protocols*, Methods in Molecular Biology, vol. 2056,
https://doi.org/10.1007/978-1-4939-9784-8_17, © Springer Science+Business Media, LLC, part of Springer Nature 2020

incurable [2]. Programmable nucleases, such as CRISPR-Cas9, are currently the most promising tools for targeted genome modifications in a variety of cell types and organisms. RNA-guided Cas9 endonucleases generate double-strand breaks (DSBs), which are repaired by two major mechanisms: nonhomologous end joining (NHEJ) and homology-directed repair (HDR) [3]. NHEJ is prone to errors resulting in indel (*in*sertion/*del*etion) mutations in target sites that may cause frameshifts and/or premature STOP codon generation. Therefore, this method can be used to knock out the expression of protein-coding genes. The more accurate but less frequently used HDR enables precise DNA modifications. In this strategy, exogenously delivered oligodeoxynucleotide (ODN) or plasmid DNA containing homology arms serves as a template for DNA synthesis and faithful repair. To increase the specificity of genome editing, one of two cleavage domains of Cas9 is mutated to form a nickase (Cas9n) that cuts only one DNA strand. DSBs are induced by using a pair of single-guide RNA (sgRNA)-Cas9n complexes targeting opposite DNA strands. This modification reduces off-target activity up to 1500-fold [4, 5] and increases the number of HDR events compared to those of wild-type Cas9 (wtCas9) [6, 7].

Several genome-editing approaches have been applied in experimental therapy for HD [8]. In an allele-selective strategy, single-nucleotide polymorphisms (SNPs) were used as a target for CRISPR-Cas9 to cut and inactivate only the mutant allele [9, 10]. In another nonallele selective strategy, *HTT* inactivation was achieved by excision of a DNA fragment containing CAG repeats by a pair of sgRNAs and wtCas9 [11].

Recently, we developed a similar strategy for *HTT* knock out by CAG repeat excision [12]. In contrast to the previous study, we use Cas9 nickase, which is more precise than wtCas9. Here, we describe a protocol for *HTT* gene knock out with this strategy. A pair of sgRNAs (sgRNA1 and sgRNA4) is designed to target sequences upstream and downstream of the CAG repeats and induce double cuts on the opposite DNA strands (Fig. 1a). Plasmid-encoded sgRNAs coexpressed with Cas9n and a GFP marker are electroporated into HD patient-derived fibroblasts containing different numbers of CAG repeat tracts (Fig. 1b). GFP-positive cells are sorted by FACS prior to DNA, RNA and protein

Fig. 1 (continued) performed in the fibroblasts (**b**) and HEK293T cells (**c**). Both cell lines were electroporated with plasmids encoding the Cas9n protein, sgRNA1 and sgRNA4 pair. Additionally, HEK293T cells were electroporated with a 179 nt ssODN as the HDR template. Two days after electroporation, fibroblast cells were sorted by FACS based on GFP expression and cultured until they reached 80% confluence. DNA, RNA and proteins were isolated and used for PCR, Sanger sequencing, qRT-PCR and Western blot analyses, respectively. After electroporation, the HEK293T cells were plated in a 150 mm cell dish and cultured until single cells formed colonies. The colonies were separated into a 48 well plate using cloning rings. When the cells reached 80% confluence, DNA was isolated from each single-cell-derived clone. The product size and sequence was confirmed by PCR and Sanger sequencing

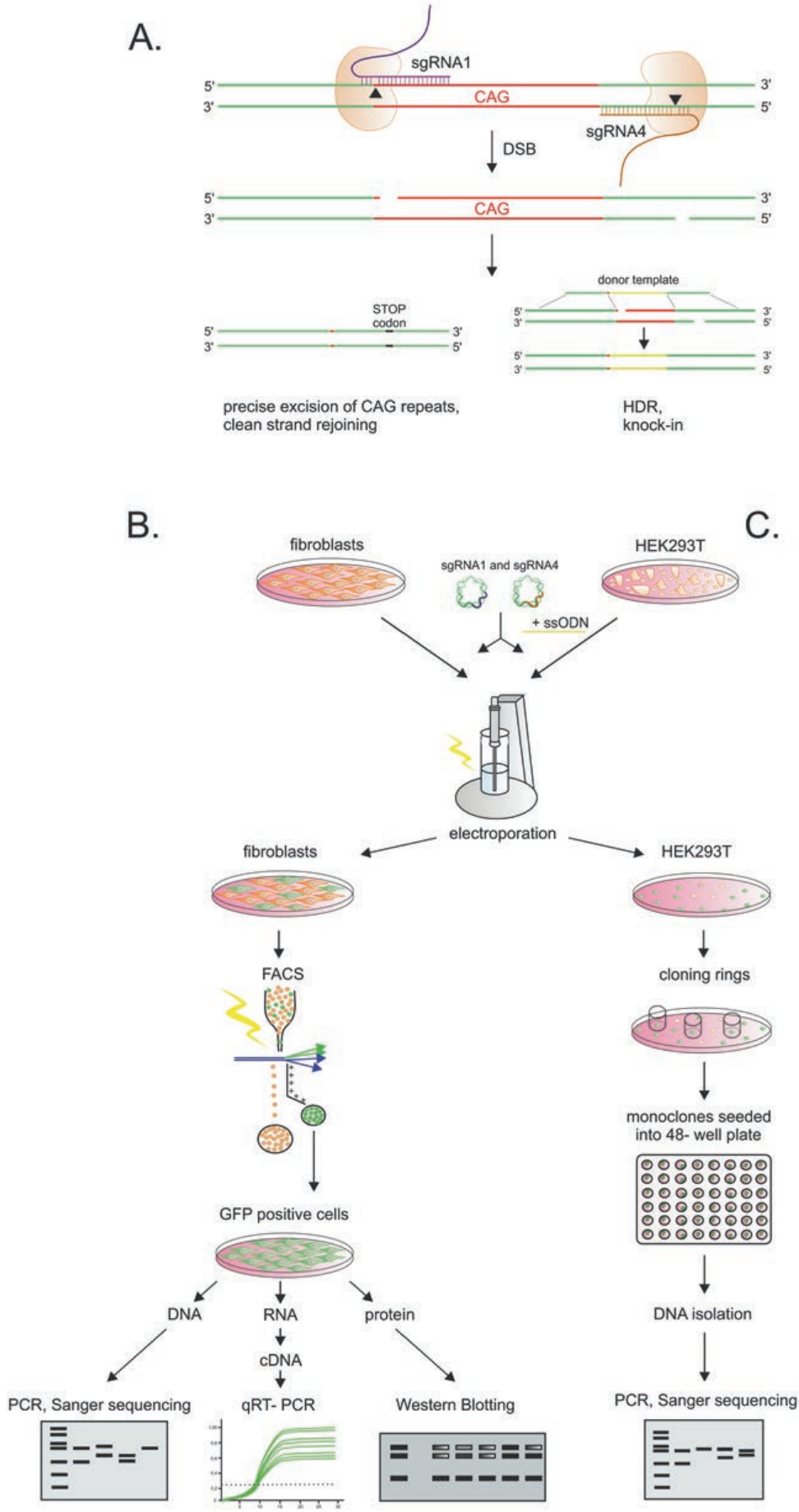


Fig. 1 A strategy utilizing a pair of sgRNAs and the Cas9n protein for precise excision of the CAG repeat tract (knock-out) or insertion of a new sequence by HDR (knockin), respectively (**a**). Schematic representation of procedures

isolation. Precise excision of CAG repeats and premature stop codon generation in the DNA and mRNA are confirmed by PCR and Sanger sequencing. The HTT protein level should be efficiently reduced by ~70% and confirmed by Western blotting.

The above strategy may also be useful for the generation of new isogenic models of HD. Therefore, in a modified version of the protocol, HEK293T cells are transfected with a pair of plasmids encoding Cas9n, sgRNA1 and sgRNA4 together with the donor template (e.g., ssODN) (Fig. 1c). The presence of a new sequence in the *HTT locus* resulting from HDR is confirmed by PCR and Sanger sequencing analysis of single-cell-derived clones.

2 Materials

2.1 Cloning

1. Plasmid: pSpCas9n(BB)-2A-GFP (D10A nickase mutant; Addgene, Cambridge, MA, USA, ID: PX461).
2. DNA oligonucleotides for sgRNA construction.
3. U6-Fwd primer (Table 1).
4. Restriction enzyme: FastDigest BpiI.
5. T4 DNA ligase, 20 U/μL.
6. 10× T4 DNA ligase buffer.
7. DNase- and RNase-free water.
8. Ampicillin.
9. LB medium.

Table 1
List of primers used

Name	Sequence
U6-Fwd	GAGGGCCTATTTCCCATGATTCC
HD1F	CCGCTCAGGTTCTGCTTTTA
HD1R	GGCTGAGGCAGCAGCGGCTG
HD_F	CGACAGCGAGTCAGTGATTG
HD_R	ACCACTCTGGCTTCACAAGG
GAPDH_F	GAAGGTGAAGGTCGGAGTC
GAPDH_R	GAAGATGGTGATGGGATTTC
sgRNA1s	CACCGCTGCTGCTGCTGCTGCTGGA
sgRNA1a	AAACTCCAGCAGCAGCAGCAGCAGC
sgRNA4s	CACCGGCTTCCTCAGCCGCCGCCGC
sgRNA4a	AAACGCGGCGGCGGCTGAGGAAGCC

10. LB agar medium.
11. ChemiComp GT116 *E. coli* (InvivoGen, San Diego, CA, USA).
12. Gene JET Plasmid Miniprep Kit (Thermo Fisher Scientific).
13. GenElute HP Endotoxin-Free Plasmid Maxiprep Kit.
14. Petri dishes, 100 mm.
15. PCR tubes.
16. Ice.
17. Fridge.
18. Thermocycler.
19. Microbiological incubator at 37 °C.
20. Bacterial incubator shaker at 37 °C.

2.2 Cell Culture

1. Patient-derived fibroblasts: GMO4208, 21/44 CAG in the *HTT* gene; GMO4281, 17/68 CAG in the *HTT* gene; and GMO9197, 21/151 CAG in the *HTT* gene (Coriell Cell Repositories, Camden, NJ, USA).
2. HEK293T cells (ATCC).
3. Minimal essential medium (MEM).
4. Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS, L-glutamine, and antibiotics.
5. Phosphate-buffered saline (PBS).
6. 10× trypsin–EDTA solution.
7. 0.4% Trypan blue solution.
8. TC20 automated cell counter and counting slides (Bio-Rad).
9. Tissue culture plates, pipettes, microtubes, 50 mL and 5 mL polypropylene tubes.
10. 5% CO_2 humidified incubator.
11. Water bath.
12. Microscope.
13. Centrifuge.

2.3 Electroporation

1. SpCas9n(BB)-2A-GFP (D10A nickase mutant) plasmid with cloned oligos corresponding to sgRNA1 and sgRNA4.
2. Single-stranded oligodeoxynucleotide (ssODN) as a donor template for HDR.
3. Neon transfection system: E and E2 buffers, R buffer, and 100 μL and 10 μL tips (Invitrogen).
4. 50 mg/mL Normocin.
5. 60-mm and 150-mm cell culture dishes.

2.4 Cell Sorting (Fibroblasts)

1. BD FACSAria III sorter (BD Biosciences).
2. FACSDIVA software (BD Biosciences).
3. Cell culture medium with 50% FBS.
4. FACS polystyrene tubes, 5 mL.

2.5 Single-Cell-Derived Clones (HEK293T)

1. Sterile cloning rings.
2. Cell culture plates, 48 wells.
3. Silicone grease and tweezers.

2.6 Genomic DNA Extraction

1. DNA isolation kit (Norgen, Thorold, ON, Canada).
2. 0.5× Direct-Lyse buffer: 10 mM Tris pH 8.0, 2.5 mM EDTA, 0.2 M NaCl, 0.15% SDS, and 0.3% Tween-20 [13].
3. Spectrophotometer DS-11 (DeNovix).

2.7 Polymerase Chain Reaction (PCR)

1. Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific).
2. Target-specific primers: HD1F and HD1R (Table 1).
3. DNase- and RNase-free water.
4. Any kit for purification of PCR products.
5. Thermal cycler.
6. PCR tubes.
7. Ice.

2.8 Gel Electrophoresis and Gel Extraction

1. Agarose powder.
2. TBE electrophoresis buffer.
3. Ethidium bromide.
4. Gel cast, combs, and gel electrophoresis tanks.
5. Gel imaging system, G:BOX (Syngene).
6. UV transilluminator.
7. Scalpel.
8. Gel extraction kit (Thermo Fisher Scientific or equivalent provider).

2.9 RNA Isolation

1. Total RNA isolation kit, Direct-zol RNA MiniPrep (Zymo Research, Irvine, CA, USA).

2.10 cDNA Synthesis

1. Superscript III Reverse Transcriptase (Life Technologies).
2. Random hexamer primers.
3. dNTPs.
4. Ribonuclease inhibitor, RNaseOUT (Invitrogen).
5. DNase- and RNase-free water.

6. Thermal cycler.
7. PCR tubes.
8. Ice.

2.11 Quantitative Real-Time PCR

1. SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad).
2. DNase- and RNase-free water.
3. Target-specific primers: HD_F, HD_R, GAPDH_F and GAPDH_R (Table 1).
4. 96-well plates.
5. CFX Connect™ Real-Time PCR Detection System (Bio-Rad).
6. CFX Manager software (Bio-Rad).

2.12 Protein Isolation

1. PB buffer: 60 mM Tris base, 2% SDS, 10% sucrose, and 2 mM PMSF.
2. PCR tubes.
3. Thermocycler.
4. Spectrophotometer.

2.13 Western Blot Analysis

1. Tris–acetate SDS-polyacrylamide gel (3–8%, NuPAGE™, Invitrogen).
2. Tris–acetate SDS running buffer (Novex, Carlsbad, CA, USA).
3. Milk (nonfat dry milk).
4. 10× PBS: 1 tablet in 100 mL (BioShop, Burlington, ON, CA).
5. PBST: PBS, 0.1% Tween 20.
6. Transfer buffer: 25 mM Tris–HCl pH 8.3, 192 mM glycine, 20% methanol, and 0.1% SDS.
7. Primary antibodies: anti-plectin, 1:1000, ab83497 (Abcam, Cambridge, UK) and anti-huntingtin, 1:1000, MAB2166 (Millipore, Burlington, MA, USA).
8. Secondary antibodies: anti-mouse HRP conjugate, 1:2000, A9917 (Sigma-Aldrich) and anti-rabbit HRP conjugate, 1:2000, 711-035-152 (Jackson ImmunoResearch, West Grove, PA, USA).
9. Western blotting device.
10. Western blotting detection solution (WesternBright™ Quantum Kit).
11. Gel imaging system (G:BOX).
12. Gel-Pro Analyzer (Media Cybernetics).

3 Methods

3.1 Cloning a sgRNA into the pSpCas9n(BB) -2A-GFP Vector for Coexpression with Cas9n

1. Digest the pSpCas9n(BB)-2A-GFP plasmid with BpiI enzyme according to the manufacturer's instructions (*see* **Notes 1** and **2**).
2. Run the digested products in a 1% agarose gel, and extract the linearized plasmid using a gel extraction kit (*see* **Note 3**).
3. Measure the DNA concentration using a spectrophotometer.
4. Construct complementary sense (s) and antisense (a) oligonucleotides that correspond to the desired sgRNAs (*see* **Notes 4** and **5**): sgRNA1s, sgRNA1a, sgRNA4s and sgRNA4a (Table 1).
5. Prepare a 100 μM solution of each oligo. Anneal the two complementary oligonucleotides. Assemble the annealing mixture as follows: 1 μL of sense oligo, 1 μL of antisense oligo and 8 μL of 1× DNA annealing buffer. Heat the mixture at 95 °C for 5 min, and then anneal at room temperature for 1 h. Dilute the annealed duplexes 1:100 in RNase- and DNase-free water.
6. Ligate the sgRNA duplex into the linearized pSpCas9n(BB)-2A-GFP plasmid. Set up a ligation reaction as follows: 50 ng of linearized plasmid, 2.5 μL of diluted oligonucleotides, 1 μL of 10× T4 ligase buffer, 0.5 μL of T4 DNA Ligase, and up to 10 μL of H_2O. Incubate the reaction at 4 °C overnight (*see* **Note 6**).

3.2 Bacterial Transformation

1. Thaw ChemiComp GT116 competent cells on ice, and incubate for 5 min on ice.
2. Pipet 45 μL of competent cells into prechilled 1.5-mL tubes, and return the tubes to the ice.
3. Add 5 μL of the ligation reaction mixture to the cells.
4. Incubate the tube on ice for 30 min.
5. Incubate the tube in a 42 °C water bath for 30 s.
6. Place the tube on ice for 2 min.
7. Add 350 μL of room temperature LB medium.
8. Incubate the tubes at 37 °C for 1 h with shaking at 300 rpm.
9. Spread 200–400 μL of the above mixture onto LB agar plates containing 100 μg/mL ampicillin.
10. Incubate the plates at 37 °C overnight.

3.3 Identification of Bacterial Colonies

1. After overnight incubation, pick approximately two colonies, and inoculate them into 3 mL of LB medium with ampicillin (100 μg/mL) in round-bottom tubes. Incubate the growth culture in a bacterial shaker with shaking at 180 rpm overnight at 37 °C.

2. Isolate the plasmid DNA using a miniprep kit (*see* **Note 7**).
3. Inoculate ~0.5 mL of the overnight culture from point 1 into 150 mL of LB medium and 100 μg/mL ampicillin in a bacterial shaker with shaking at 180 rpm overnight at 37 °C.
4. After overnight incubation, isolate the plasmid DNA using an Endotoxin-free Maxiprep Kit according to the manufacturer's instructions.
5. Measure the concentration of the isolated plasmid by using a spectrophotometer.
6. Store the DNA at −20 °C or use immediately.

3.4 Fibroblast and HEK293 Cell Cultures

1. Grow the fibroblasts in MEM supplemented with 10% FBS, L-glutamine (2 mM) and antibiotic (100 U penicillin, 10 mg streptomycin, and 25 μg amphotericin) in a 10 cm culture plate in a 5% CO_2 incubator at 37 °C.
2. Grow the HEK293T cells in DMEM supplemented with 10% FBS, L-glutamine and antibiotics in a 10 cm dish in an incubator supplied with 5% CO_2.
3. Change the medium every 2 days, and harvest the cells when they exceed 80% confluence.
4. Wash the cells with 10 mL of PBS.
5. Harvest the cells using a 1× trypsin–EDTA solution, and the incubate the cells at 37 °C for 3–5 min in an incubator supplied with 5% CO_2.
6. Examine the cells under a microscope to ensure that they have completely detached from the plates, and add 5 mL of complete medium to neutralize the trypsin.
7. Add 7 μL of cells suspended in medium to 7 μL of trypan blue, and then count the cells using a cell counter.

3.5 Electroporation

1. Electroporate the cells with the Neon™ Transfection System (*see* **Note 8**). Harvest 1×10^6 to 5×10^5 fibroblast cells, resuspend the cells in PBS, and electroporate the cells with 10 μg of plasmid DNA (5 μg of each plasmid containing sgRNA1 and sgRNA4) (*see* **Note 9**) in 100 μL tips according to the manufacturer's instructions. Use the following parameters: 1350 V, 30 ms, and 1 pulse.
2. Electroporate 5×10^4 HEK293T cells suspended in R buffer with 125 ng of plasmid (62.5 ng of each plasmid containing sgRNA1 and sgRNA4) and 0.5 μL of 10 μM ssODN in 10 μL tips using the following parameters: 1150 V, 20 ms, and 2 pulses (*see* **Note 10**).

3. Seed the fibroblasts after electroporation into a 60 mm cell culture plate with MEM supplemented with 10% FBS and L-glutamine without antibiotics.
4. Plate the HEK293T cells in a 150 mm plate (~1000 cells per plate), and grow the cells until they form single cell clones.
5. Four hours after electroporation, add 100 μg/mL Normocin to the medium.
6. Check the transfection efficiency 24 h after electroporation. The percentage of fluorescent cells (GFP expression) can by estimated by using a fluorescence microscope (*see* **Note 11**).

3.6 Fluorescence-Activated Cell Sorting (FACS) (*See* Note 12)

1. Sort the cells 48 h post-electroporation.
2. Wash the cells with PBS, harvest them using 1× trypsin–EDTA solution, add complete medium and centrifuge.
3. Suspend the cell pellet in 1 mL of PBS. Set the configuration of the flow cytometer as follows: a 100-m nozzle and 20 psi (0.138 MPa) of sheath fluid pressure.
4. Collect the sorted cells into tubes containing 2.5 mL of medium supplemented with 50% FBS.
5. Analyze the data with FACSDIVA software.
6. Seed the cells into a 6-well plate, add 100 μg/mL Normocin to the medium and maintain them until they reach confluence. Then, collect the cells for genomic DNA extraction (fibroblasts), or seed them into a 150-mm plate (HEK293T cells).

3.7 Single-Cell-Derived Clones from HEK293T Cells

1. Gently wash cells seeded into a 150 mm plate with 10 mL of PBS.
2. Mark individual colonies.
3. Using sterile curved forceps, center the cloning ring over a cell colony (*see* **Note 13**).
4. Add 20 μL of trypsin–EDTA solution directly into the cylinder, and incubate at 37 °C for 2 min in a 5% CO_2 incubator. Monitor the colony under a microscope.
5. Neutralize the trypsin by adding 80 μL of DMEM.
6. Transfer the detached cells from a single colony to a 48-well plate.
7. Grow the cells in a 48-well plate until they reach 60%–80% confluence (*see* **Note 14**).

3.8 DNA Extraction

1. Extract DNA from the fibroblasts using a Cell and Tissue DNA Isolation Kit according to the manufacturer's instructions. Measure the DNA concentration using a spectrophotometer.
2. Isolate DNA from individual HEK293T colonies using 0.5× Direct-Lyse buffer as described by Ramlee et al. [13].

3.9 PCR and Gel Electrophoresis

1. Mix 5 μL of 2× Phusion Master Mix, 1 μL of primers (10 μM mixture of the forward and reverse primers), 100 ng of DNA template and H_2O to a final volume of 10 μL (*see* **Note 15**). Incubate the mixture in a thermal cycler at 98 °C for 3 min, followed by 12 cycles at 98 °C for 15 s and 60 °C for 15 s, an additional 21 cycles at 98 °C for 15 s and 62 °C for 15 s, and a final incubation at 72 °C for 5 min.
2. Run the PCR products with gel loading dye on a 1.3% agarose gel. Use an appropriate DNA ladder. Visualize the products using a gel imaging system.
3. Purify the PCR products.
4. Confirm the sequence of the purified PCR products using Sanger sequencing with the forward primer HD1F (Table 1) (*see* **Note 16** and **17**).

3.10 RNA Isolation and cDNA Synthesis

1. Remove the medium from the cell culture plate, and wash the cells once with PBS.
2. Add an appropriate amount of TRIzol reagent directly into a cell culture dish, and isolate the RNA using a Direct-zol RNA MiniPrep kit according to the manufacturer's instructions. Use the DNase treatment step.
3. To elute the RNA, add DNase- and RNase-free water or TE buffer, and measure the RNA concentration using a spectrophotometer.
4. To perform cDNA synthesis, mix 0.5–1 μg of RNA, 1 μL of 10 μM dNTPs, and 0.25 μL of random primers (500 μg/mL).
5. Incubate the samples at 65 °C for 5 min, and then place them on ice for 1 min. Perform a quick spin, and prepare a new mixture containing 4 μL of 5× First Strand Buffer, 1 μL of 0.1 M DTT, 1 μL of RNase OUT and 1 μL of Superscript III RT.
6. Mix by pipetting, and incubate at 25 °C for 5 min, 55 °C for 60 min and 70 °C for 15 min. Store the cDNA at −20 °C.

3.11 Quantitative Real-Time PCR Analysis

1. Use 700 ng of complementary cDNA for quantitative polymerase chain reaction (qPCR) using SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad). Prepare three replicates of each sample with HTT and GAPDH primers.
2. Mix 5 μL of SsoAdvanced™ Universal SYBR® Green Supermix with 1 μL of 10 μM primers (HTT or GAPDH), 3 μL of H_2O and 1 μL of cDNA (*see* **Note 18**).
3. Incubate the samples in the CFX Connect™ Real-Time PCR Detection System under the following conditions: incubation at 95 °C for 30 s, followed by 40 cycles at 95 °C for 15 s and 60 °C for 30 s, and then incubation at 65 °C–95 °C (delta 0.5 °C per 5 s).
4. Analyze the data by using CFX Manager software.

3.12 Protein Isolation

1. Remove the medium from the cells described above, and wash the cells with PBS.
2. Use trypsin–EDTA solution to detach the cells from the dish.
3. Neutralize the trypsin–EDTA solution by using MEM containing 10% FBS.
4. Centrifuge the cells for 5 min at 268 × *g*. Remove the supernatant from the cell pellet.
5. Use PB buffer to isolate total proteins. Add approximately 30 μL of PB buffer to the cell pellet, and denature the proteins in a thermal cycler for 5–10 min at 90 °C.
6. Measure the protein concentration by using a spectrophotometer.
7. Store the isolated proteins at −20 °C, or use immediately for Western blot analysis.

3.13 Western Blotting

1. Mix 30 μg of protein with a sample loading buffer containing 2-mercaptoethanol, and boil for 5 min at 95 °C.
2. Load 12 μL of HiMark™ Prestained Protein Standard and ~15 μL of each sample onto a Tris–acetate SDS-polyacrylamide gel (3–8%), and run the electrophoresis in Tris–acetate SDS running buffer at 170 V and 4 °C.
3. After electrophoresis, equilibrate the SDS-PAGE gels in transfer buffer for 15 min at room temperature.
4. Cut the membrane roughly to the size of the SDS-PAGE gel. Remember the approximate size of the huntingtin (350 kDa) and plectin (500 kDa) proteins. Place the nitrocellulose membrane immediately in transfer buffer.
5. Cut eight pieces of Whatman 3-mm chromatography paper (four pieces per site) to the size of the gel, and wet each piece in transfer buffer.
6. Make a gel sandwich in the following order: (top) one sponge, four wet pieces of Whatman paper, one wet layer of 3-mm nitrocellulose membrane, SDS gel, four wet pieces of Whatman paper and one sponge (bottom). Take care to remove air bubbles by rolling each layer with a clean roller.
7. Electrophoretically transfer the proteins to the membrane at 30 V overnight.
8. Check the transfer efficiency by Ponceau staining. Wash the membrane with PBST. During this step, the membrane may be cut into two pieces containing the huntingtin or plectin protein bands.
9. Place the membrane into a petri dish, and block nonspecific binding using 5% milk dissolved in PBST for 1 h.

10. Prepare the primary and secondary antibody solutions in milk (*see* **Note 19**).
11. Decant the milk, add the primary antibody solution to the membrane in a clean Petri dish, and incubate for 2 h at room temperature in a 3D blot mixer. The membrane should be completely covered by the primary antibody solution. If 10 mL is not sufficient to completely cover the membrane, larger volumes may be used.
12. Decant the primary antibody solution, and wash the membrane three times for 5–20 min with 50 mL of PBST at room temperature in a 3D blot mixer.
13. Add the secondary antibody solution, and incubate the membrane for 1 h at room temperature in a 3D blot mixer.
14. Remove the antibody solution, and wash the membrane three times for 5–20 min with 50 mL of PBST at room temperature in a 3D blot mixer.
15. Prepare the Western blotting detection solution according to the manufacturer's instructions, and add it to the membrane, ensuring that the entire membrane is covered with a thin layer of liquid. Incubate for 5 min in the dark.
16. Put the membrane in a Petri dish cover to dry, and place directly into the G-BOX.
17. Perform several exposures of the membrane such that each band of interest and the marker have the maximum signal. Typically, reference proteins such as plectin give a strong signal and require separation from the huntingtin band.
18. Perform quantified analysis of the bands using Gel-Pro Analyzer

4 Notes

1. The pSpCas9n(BB)-2A-GFP plasmid also expresses the GFP protein, which reveals the transfection efficiency.
2. The Cas9n protein and target-specific sgRNA are encoded in the same vector. For double nicking application, use a 1:1 ratio of plasmids expressing sgRNA1 and sgRNA4.
3. After digestion of the plasmid, ensure that the digestion was completed properly. Check the sample after digestion and without digestion on a 1% agarose gel with a 1 kb DNA ladder.
4. To design other sgRNAs, use CRISPOR (http://crispor.tefor.net/) [14] or other software as described in Cui et al. [15]. This tool identifies suitable target sites and predicts potential

off-targets for each intended part of the genome. Order the necessary oligos as specified by the online tool.

5. An extra "G" has to be added at the 5′ end of the sgRNA when the guide sequence does not begin with "G". The U6 RNA polymerase III promoter that expresses the sgRNA prefers a guanine nucleotide as the first base of its transcript. This trick facilitates sgRNA expression.
6. To properly conduct ligation, purify the digested plasmid, and if necessary, dephosphorylate the cut ends. The digested bands can be isolated from the agarose gel and dephosphorylated using a commercial kit. This step may be necessary to avoid self-ligation and circularization of the plasmid DNA.
7. Verify that the sgRNA oligos were cloned into the plasmid by sequencing from the U6 promoter using the U6-Fwd primer or performing a PCR with correctly designed primers. The primers should be located near an appropriate restriction site in the plasmid sequence.
8. For transfection of HEK293T and fibroblast cells, we recommend nucleofection (Neon transfection system). The Neon™ website contains many protocols and conditions suitable for various cell lines.
9. Vectors delivered into cells via electroporation have to be highly concentrated (concentration between 1 and 5 μg/μL) and free of bacterial toxins. The DNA amount should not exceed 10% of the total reaction volume because this amount influences the transfection efficiency and cell viability after nucleofection.
10. The donor template can be a ssODN, plasmid or PCR product that contains the sequence necessary for strand synthesis during HDR. The ssODN is comprised of two homology arms and the desired sequence. The homology arms need to be complementary to the sequences around the DSB, and their lengths depend on the size of the inserted sequence (usually approximately 30–70 nt). In our experiments, a 179 nt ssODN was used that consisted of two 60 nt homology arms and the 59 nt GFP sequence.
11. The electroporation efficiency depends on the vector size. If the plasmids are more than 8 kb, the transfection efficiency tends to be lower [16].
12. Sort the cells into collection tubes with appropriate medium and a high concentration of FBS. Cells are weak after sorting and need more supplements for recovery.
13. Methods used to generate a monoclonal cell line include serial dilutions or single-cell sorting (one cell per well) into a 96-well plate by using a FACS device. Cells can also be seeded into a

150 mm dish and detached after they reach the appropriate confluency by using cloning rings. Colonies have to be separated from each other to avoid cross-contamination.

14. One or two weeks may be required to form colonies from a single cell. Change the medium supplemented with FBS and antibiotics every 2 days. For better cell growth, make conditioned medium (1:1 ratio of filtered medium from HEK293T cells and new complete DMEM).
15. Using high-fidelity polymerases is important to minimize errors during PCR amplification. Design PCR primers outside the sequence complementary to the homology arms to avoid detection of residual donor template.
16. Modified DNA (after excision of the CAG repeats from the *HTT* gene) is clearly visible on a gel as a shorter PCR product. The products can by purified by using any gel extraction kit. Only purified PCR products can be used for Sanger sequencing.
17. Generally, indel mutations can be detected by mismatch cleavage assays, such as T7E1, a Surveyor assay or other methods, such as ddPCR (droplet digital PCR), IDAA (indel detection by amplicon analysis), deep sequencing, or qEva-CRISPR [17]. However, PCR amplification of the *HTT* gene region, including polymorphic CAG repeats, and the subsequent T7E1 analysis result in the generation of multiple bands in both the treated and untreated control cells. Therefore, a determination of the exact indel frequency in this polymorphic, highly repetitive gene region is difficult using methods based on heteroduplex recognition. We suggest using the qEva-CRISPR assay, which allows for successful analysis of targets located in "difficult" genomic regions [17].
18. Prepare and aliquot the real-time PCR reaction mixture in the dark. Light as well as repeated cycles of thawing and freezing inactivate fluorescent dyes.
19. Dilute the primary antibodies (anti-huntingtin and anti-plectin) and secondary antibodies (anti-mouse HRP conjugate and anti-rabbit HRP conjugate) in PBS/0.1% Tween-20 buffer containing 2.5% nonfat milk to final dilutions of 1:1000 and 1:2000, respectively.

Acknowledgments

This work was supported by a grant from the National Science Center (2015/18/E/NZ2/00678) and from the quality-promoting subsidy under the Leading National Research Center (KNOW) program for 2014–2018.

References

1. Bates GP, Dorsey R, Gusella JF et al (2015) Huntington disease. Nat Rev Dis Primers 1:15005. https://doi.org/10.1038/nrdp.2015.5
2. Wild EJ, Tabrizi SJ (2017) Therapies targeting DNA and RNA in Huntington's disease. Lancet Neurol 16:837–847. https://doi.org/10.1016/S1474-4422(17)30280-6
3. Shibata A, Steinlage M, Barton O et al (2017) DNA double-strand break resection occurs during non-homologous end joining in G1 but is distinct from resection during homologous recombination. Mol Cell 65:671–684.e5. https://doi.org/10.1016/j.molcel.2016.12.016
4. Cho SW, Kim S, Kim Y et al (2014) Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res 24:132–141. https://doi.org/10.1101/gr.162339.113
5. Ran FA, Hsu PD, Lin CY et al (2013) Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. Cell 154:1380–1389. https://doi.org/10.1016/j.cell.2013.08.021
6. Kocher T, Peking P, Klausegger A et al (2017) Cut and paste: efficient homology-directed repair of a dominant negative KRT14 Mutation via CRISPR/Cas9 nickases. Mol Ther 25:2585–2598. https://doi.org/10.1016/j.ymthe.2017.08.015
7. Eggenschwiler R, Moslem M, Fráguas MS et al (2016) Improved bi-allelic modification of a transcriptionally silent locus in patient-derived iPSC by Cas9 nickase. Nat Publ Gr:1–14. https://doi.org/10.1038/srep38198
8. Mosbach V, Poggi L, Richard G-F (2019) Trinucleotide repeat instability during double-strand break repair: from mechanisms to gene therapy. Curr Genet 65:17–28. https://doi.org/10.1007/s00294-018-0865-1
9. Shin JW, Kim K-H, Chao MJ et al (2016) Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9. Hum Mol Genet 25(20):4566–4576. https://doi.org/10.1093/hmg/ddw286
10. Monteys AM, Ebanks SA, Keiser MS, Davidson BL (2017) CRISPR/Cas9 editing of the mutant Huntingtin Allele in vitro and in vivo. Mol Ther 25:12–23. https://doi.org/10.1016/j.ymthe.2016.11.010
11. Yang S, Chang R, Yang H et al (2017) CRISPR/Cas9-mediated gene editing ameliorates neurotoxicity in mouse model of Huntington's disease. J Clin Invest 127:2719–2724. https://doi.org/10.1172/JCI92087
12. Dabrowska M, Juzwa W, Krzyzosiak WJ, Olejniczak M (2018) Precise excision of the CAG tract from the Huntingtin gene by Cas9 nickases. Front Neurosci 12:75. https://doi.org/10.3389/FNINS.2018.00075
13. Ramlee MK, Yan T, Cheung AMS et al (2015) High-throughput genotyping of CRISPR/Cas9-mediated mutants using fluorescent PCR-capillary gel electrophoresis. Sci Rep 5:15587. https://doi.org/10.1038/srep15587
14. Haeussler M, Schönig K, Eckert H et al (2016) Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol 17:148. https://doi.org/10.1186/s13059-016-1012-2
15. Cui Y, Xu J, Cheng M et al (2018) Review of CRISPR/Cas9 sgRNA design tools. Interdiscip Sci Comput Life Sci 10:455–465. https://doi.org/10.1007/s12539-018-0298-z
16. Yumlu S, Stumm J, Bashir S et al (2017) Gene editing and clonal isolation of human induced pluripotent stem cells using CRISPR/Cas9. Methods 121–122:29–44. https://doi.org/10.1016/j.ymeth.2017.05.009
17. Dabrowska M, Czubak K, Juzwa W et al (2018) qEva-CRISPR: a method for quantitative evaluation of CRISPR/Cas-mediated genome editing in target and off-target sites. Nucleic Acids Res 46:e101. https://doi.org/10.1093/nar/gky505

Published online 7 June 2018

Nucleic Acids Research, 2018, Vol. 46, No. 17 ***e101***
doi: 10.1093/nar/gky505

qEva-CRISPR: a method for quantitative evaluation of CRISPR/Cas-mediated genome editing in target and off-target sites

Magdalena Dabrowska[1,†], Karol Czubak[2,†], Wojciech Juzwa[3], Wlodzimierz J. Krzyzosiak[4], Marta Olejniczak[1,*] and Piotr Kozlowski[2,*]

[1]Department of Genome Engineering, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland, [2]Department of Molecular Genetics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland, [3]Department of Biotechnology and Food Microbiology, Poznan University of Life Sciences, Wojska Polskiego 48, 60-627 Poznan, Poland and [4]Department of Molecular Biomedicine, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland

Received April 04, 2018; Editorial Decision May 17, 2018; Accepted May 23, 2018

ABSTRACT

Genome editing technology based on engineered nucleases has been increasingly applied for targeted modification of genes in a variety of cell types and organisms. However, the methods currently used for evaluating the editing efficiency still suffer from many limitations, including preferential detection of some mutation types, sensitivity to polymorphisms that hamper mismatch detection, lack of multiplex capability, or sensitivity to assay conditions. Here, we describe qEva-CRISPR, a new quantitative method that overcomes these limitations and allows simultaneous (multiplex) analysis of CRISPR/Cas9-induced modifications in a target and the corresponding off-targets or in several different targets. We demonstrate all of the advantages of the qEva-CRISPR method using a number of sgRNAs targeting the *TP53*, *VEGFA*, *CCR5*, *EMX1* and *HTT* genes in different cell lines and under different experimental conditions. Unlike other methods, qEva-CRISPR detects all types of mutations, including point mutations and large deletions, and its sensitivity does not depend on the mutation type. Moreover, this approach allows for successful analysis of targets located in 'difficult' genomic regions. In conclusion, qEva-CRISPR may become a method of choice for unbiased sgRNA screening to evaluate experimental conditions that affect genome editing or to distinguish homology-directed repair from non-homologous end joining.

INTRODUCTION

Genome-editing technology is widely used to inactivate or modify specific genes in functional studies or in therapeutic approaches. The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 system (1–4) recently became a major genome editing tool that has replaced previously developed zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) (5,6). In the CRISPR/Cas9 system, single guide RNA (sgRNA) is used to guide the Cas9 nuclease to target DNA containing the protospacer adjacent motif (PAM), which is 5′-NGG-3′ for *Streptococcus pyogenes* Cas9. Double-strand breaks (DSBs) generated by Cas9 at ∼3 bp upstream from PAM are mainly repaired by error-prone non-homologous end joining (NHEJ), which results in a variety of scar mutations, most of which are insertion/deletion (INDEL) frameshift mutations leading to premature translation termination and transcript degradation by nonsense-mediated decay (NMD). Alternatively, the homology-directed repair (HDR) mechanism can repair the break precisely using a DNA repair template. The mechanism of CRISPR/Cas9-mediated genome editing has been recently described in detail in several excellent review articles, e.g. (7). There are four possible results of target gene editing in a single diploid cell: no mutation, heterozygous mutation (only one allele is mutated), homozygous mutation (the same mutation in both alleles), and biallelic mutation (different mutations on both alleles). Despite great progress in sgRNA design algorithms (8–10), the efficiency of a specific DSB induction within the target sequence is still difficult to predict. Additionally, unspecific targeting of other genomic regions (off-targets) is difficult to avoid and therefore remains one of the most im-

*To whom correspondence should be addressed. Tel: +48 61 8528503; Fax: +48 61 8520532; Email: kozlowp@ibch.poznan.pl
Correspondence may also be addressed to Marta Olejniczak. Tel: +48 61 8528503; Fax: +48 61 8520532; Email: marta.olejniczak@ibch.poznan.pl
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

© The Author(s) 2018. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

portant challenges of genome editing approaches, especially in the context of their clinical applications [reviewed in (11)].

Several methods have been developed to evaluate the activity of sgRNAs and frequency of INDEL mutations; however, all of them have their specific limitations (12). Most methods, including mismatch cleavage assays, high-resolution melting analysis (HRMA), and heteroduplex mobility, are based on cleavage or modified migration of the heteroduplexes formed by mutated and wild-type DNA strands (6,13–15). These methods are widely used for preliminary screening of sgRNA activity due to their simplicity, low cost, and requirements for basic laboratory equipment. The most popular of these techniques utilize T7 endonuclease 1 (T7E1) or Surveyor nuclease (Transgenomic, Inc., USA) to cleave mismatches formed between modified and unmodified DNA strands (12,16). Despite these advantages, mismatch cleavage assays can overlook both single-nucleotide changes as well as larger deletions. They also cannot detect homozygous mutations and are not suitable for analyses of polymorphic loci (17). Other INDEL detection methods include restriction fragment length polymorphism (RFLP) (5,18), loss of a primer binding site (19), analysis of a PCR product length polymorphism (20), and decomposition of Sanger sequencing reads by TIDE (21,22) and CRISPR-GA (23). Alternative methods have also been proposed (24,25). Unlike heteroduplex-utilizing assays, these methods allow for detection of homozygous mutant clones. Most of the abovementioned methods take advantage of the principles of mutation detection methods that were developed and commonly used in the 1990s and are affected by their limitations, of which the most important are as follows: limited sensitivity (usually ∼80%), the confounding effects of common SNPs that frequently occur in the vicinity of the site of interest, and an inability to detect larger (>200) deletions/duplications. Most importantly, many of these methods were not designed to be quantitative and have limited or no multiplexing capabilities.

Functional screening of candidate sgRNAs rarely includes a more challenging analysis of off-targets. Although bioinformatics tools are helpful to predict potential sgRNA off-targets in the genome of interest [e.g. (12,26)], the number of off-targets may depend on many factors, including the thermodynamic properties of sgRNA and actual local concentrations of sgRNA and Cas9 as well as the local ion concentrations and presence of various specific and unspecific competing or sequestrating elements. The latter factor further depends on experimental conditions, such as the cell type or sgRNA/Cas9 delivery strategy [reviewed in (11)].

Here, we describe a new method for **q**uantitative **Eva**luation of **CRISPR**/Cas9-mediated editing (qEva-CRISPR) that allows for parallel analysis of target and selected off-target sites. qEva-CRISPR is a ligation-based dosage-sensitive method. It takes advantage of the previously developed and well-validated strategy of the multiplex ligation-based probe amplification (MLPA) assay design (27,28). The strategy exclusively utilizes short oligonucleotide probes that can be easily generated via standard chemical synthesis for any target of interest. The general concept of the MLPA method, which is mostly used for detection of large deletions in disease-related genes, has been described and reviewed previously (29,30).

Using the developed qEva-CRISPR assays, we quantitatively measured the efficiency of several sgRNAs in target and off-target sites in different human cell lines and under different conditions in CRISPR/Cas9 experiments. Among the analyzed targets were those located within low-complexity sequences flanking the CAG microsatellite tract in the *huntingtin* (*HTT*) gene, for which other assays could not be designed or did not work reliably. We demonstrate the ability of qEva-CRISPR to distinguish sequences generated by NHEJ and HDR.

MATERIALS AND METHODS

Plasmids

The sgRNA sequences E4.1 and E4.2, which are specific for target sequences within the *TP53* gene, and sgRNAs specific for the *VEGFA* (VEGFA), *EMX1* (EMX1), and *CCR5* (CCR5.1, CCR5.6, and CCR5.7) genes have been described previously (31–33). To generate the Cas9_E4.1, Cas9_E4.2, Cas9_VEGFA, Cas9_EMX1, Cas9_CCR5.1, Cas9_CCR5.6 and Cas9_CCR5.7 plasmids, sense and antisense DNA strands of sgRNAs were synthesized (IBB, Warsaw, Poland), annealed to each other, and ligated into the FastDigest BsmBI (Thermo Fisher Scientific, Waltham, MA, USA) digested pSpCas9(BB)-2A-GFP (PX458) (Addgene, Cambridge, MA, USA) plasmid. Chemically competent *Escherichia coli* GT116 cells (InvivoGen, San Diego, CA, USA) were transformed with the ligated plasmids, plated onto ampicillin selection plates (100 μg/ml ampicillin) and incubated overnight at 37°C. The plasmids were isolated using the Gene JET Plasmid Miniprep kit (Thermo Fisher Scientific) and analyzed by Sanger sequencing with the U6-Fwd primer. The same strategy was used to generate plasmids encoding Cas9 and three sgRNAs specific for the *HTT* gene (34): Cas9_HTT.sg1, Cas9_HTT.sg3 and Cas9_HTT.sg4. Two *HTT* sgRNAs were also used with the nickase version of Cas9 (Cas9n; D10A mutant; pSpCas9n(BB)-2A-GFP (PX461)): Cas9n_HTT.sg1 and Cas9n_HTT.sg4. The oligonucleotide sequences are presented in Supplementary Table S1.

Cell culture and transfection

Human colon cancer cells (HCT116), human embryonic kidney cells (HEK293T), human myeloid leukemia cells (K562) and HeLa cells were grown in Dulbecco's modified Eagle's medium (Lonza; Basel, Switzerland) supplemented with 10% fetal bovine serum (Sigma-Aldrich, Saint Louis, MS, USA), antibiotics (Sigma-Aldrich) and L-glutamine (Sigma-Aldrich). HCT116 and HeLa transfections were performed using Lipofectamine LTX (Life Technologies, Carlsbad, CA, USA) according to the manufacturer's instructions. Plasmids were transfected with a concentration of 0.25 μg to 2 μg/12-well plate. The transfection efficiency (GFP positive cells) was determined by flow cytometry (BD Accuri™ C6, BD Biosciences, Franklin Lakes, NJ, USA). HEK293T were transfected using the calcium phosphate method with 10 μg of plasmid DNA or 5 μg of each plasmid from a Cas9n_HTT.sg1 and Cas9n_HTT.sg4 pair for 3×10^5 cells/plate. HCT116 and K562 cells were electroporated with the Neon™ Transfection System (Invitrogen,

Carlsbad, CA, USA). Briefly, 0.5–1 × 10^5 cells were harvested, resuspended in Buffer R and electroporated with 1 μg of plasmid DNA in 10 μl tips using the following parameters: 1130 V, 30 ms, 2 pulses or 1450 V, 10 ms, 3 pulses for HCT116 and K562 cells, respectively. In multiplex analysis, the concentration of each plasmid DNA was either 200 ng (0.6 μg in total) or 330 ng (1 μg in total). In HDR experiments, the HEK293T cells were electroporated with 125 ng of plasmid DNA (62.5 ng of each Cas9n_HTT.sg1 and Cas9n_HTT.sg4 plasmids) and 0.5 μl of 10 μM oligodeoxynucleotide (ssODN) as a donor template (IDT, Skokie, IL, USA) (Supplementary Table S1) using the following parameters: 1150 V, 20 ms, 2 pulses. The cells were cultured for 48 hours, after which genomic DNA was isolated with the Cells and Tissue DNA Isolation Kit (Norgen, Biotek Corp., Schmon Pkwy, ON, Canada) according to the manufacturer's instructions.

In vitro T7 transcription of sgRNA and RNP complex delivery

R_1s and R_1a oligonucleotides corresponding to HTT_sg.1 (IBB, Warsaw) (Supplementary Table S1) were annealed and ligated into the FastDigest *Bpi*I (Thermo Fisher Scientific) digested p31 vector, which contains a T7 RNA polymerase promoter and Cas9 gRNA scaffold sequence. Chemically competent *E. coli* GT116 cells (InvivoGen) were transformed with the ligated plasmids, plated onto ampicillin selection plates (100 μg/ml ampicillin) and incubated overnight at 37°C. The plasmids were isolated using the Gene JET Plasmid Miniprep kit (Thermo Fisher Scientific) and analyzed by Sanger sequencing with the WSF6 primer. The sgRNA expression vectors were digested with FastDigest *Dra*I (Thermo Fisher Scientific), and the sgRNA was synthesized using an AmpliScribe T7-Flash Transcription Kit (Epicentre, Madison, WI, USA). The synthesized sgRNA was purified by phenol–chloroform–isoamyl alcohol (PanReac AppliChem, Barcelona, Spain) extraction and its integrity was checked by electrophoresis on a 10% PAA/urea/1× TBE gel.

The RNP complex was produced by mixing recombinant NLS-SpCas9-NLS nuclease (VBCF Protein Technologies facility http://www.vbcf.ac.at) and *in vitro* transcribed sgRNA. Cas9 RNPs were prepared immediately before electroporation by incubating 2.5, 5 and 10 μg of Cas9 protein with 6, 12 and 24 μg of sgRNA transcript in an appropriate buffer containing 200 mM HEPES (pH 7.5), 1.5 M KCl, 5 mM DTT and 1 mM EDTA at 37°C for 10 min. HEK293T cells were electroporated with a Neon transfection system (Invitrogen) according to the manufacturer's instruction.

Fluorescence-activated cell sorting

Cells were sorted using the BD FACSAria™ III (BD Biosciences) flow cytometer (cell sorter) 48 h post-electroporation. The configuration of the flow cytometer was as follows: a 100-μm nozzle and 20 psi (0.138 MPa) of sheath fluid pressure. Cells were characterized by two non-fluorescent parameters, forward scatter (FSC) and side scatter (SSC), and one fluorescent parameter, which was green fluorescence from GFP collected using the 530/30 band-pass filter (FITC detector). Data were acquired in a four-decade logarithmic scale as area signals (FSC-A, SSC-A, and FITC-A) and analyzed with FACS DIVA software (BD Biosciences). GFP-positive cells were divided into three fractions based on the fluorescence intensity. Each fraction contained ∼1–2 × 10^4 cells that were seeded onto a 6-well plate, maintained until confluence and collected for genomic DNA extraction.

Single-cell clones

To obtain single-cell clones, the HCT116 cells were transfected with the Cas9_E4.1 or Cas9_E4.2 plasmids, sorted by FACS and plated onto 150-mm cell plates to form single-cell clones. Individual colonies were picked with the use of cloning rings and transferred into 48-well plates according to the manufacturer's instructions (Promega, Madison, WI, USA). DNA was isolated from cells using 0.5 × Direct-Lyse buffer as described by Ramlee et al (30). PCR was performed using the MK024 and MK025 primers, and single-cell clones were genotyped by Sanger sequencing with the MK024 primer (Supplementary Table S1).

T7E1 mutation detection assay

For the T7E1 mismatch assay, genomic DNA was amplified using Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific) with primers (MK024 and MK025) as described in the paper by Ramlee et al (30). The PCR amplification conditions were as follows: initial denaturation for 3 min at 95°C; 30 cycles of 95°C for 30 s, 63°C for 30 s, and 72°C for 30 s; and a final elongation at 72°C for 5 min. PCR products were purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific). Next, 400 ng of purified PCR product was used in an annealing reaction and enzymatic digestion by the T7E1 enzyme (New England Biolabs, Ipswich, MA, USA). The DNA fragments were separated on a 1.3% agarose gel and visualized by EB staining. The DNA band intensity was quantified using G:BOX (Syngene, Cambridge, UK) and analyzed with GelPro software (Media Cybernetics, Rockville, MD, USA). The INDEL frequency was estimated by comparing the digested band intensities to all bands.

The qEva-CRISPR assay generation

The qEva-CRISPR analysis was performed using the following custom-designed assays (probe mixes) with 7–12 MLPA probes: qEva_E4.1, qEva_E4.1_delA, qEva_E4.1_A/T, qEva_E4.2, qEva_HTT.sg1, qEva_HTT.sg3, qEva_HTT.sg4, qEva_HTT.edit, qEva_SNP, qEva_VEGFA, qEva_EMX1, qEva_CCR5.1, qEva_CCR5.6, qEva_CCR5.7 and qEva_multiplex. In most cases, the assay IDs corresponded to the IDs of the tested sgRNA/targets (Supplementary Table S2). Most assays consisted of one target specific (TS) probe and one to three off-target specific (OS) probes. The qEva_HTT.edit assay consisted of two probes (TS_HTT.sg1* and TS_HTT.sg4) specific for targets of HTT.sg1 and HTT.sg4 as well as one

probe (TS_HTT.HDR) specific for sequence overlapping the ssODN sequence (inserted to genome by HDR) and the genomic sequence flanking the incorporated ssODN sequence. Additionally, the 5′ half-probe of TS_HTT.sg1* and the 3′ half-probe of TS_HTT.sg4 formed a new probe (TS_HTT.sg1/4) that is specific to a new sequence created after rejoining of free-ends created after the excision of the sequence between the HTT.sg1 and HTT.sg4 cuts. In addition to the TS and OS probes, each assay consisted of up to eight control probes. The sequences, genomic positions, and detailed characteristics of the probes used in this study are presented in Supplementary Table S3. The qEva-CRISPR probes were designed according to a strategy previously proposed for MLPA probes (27,28). Three three-oligonucleotide probes (TS_HTT.HDR, TS_CCR5.1 and OS_CCR5.1) were designed according to a strategy adopted from (35,36). All probes were synthesized by IDT, Skokie, IL, USA. All reagents, except for the probe mixes, were purchased from MRC-Holland, Amsterdam, The Netherlands. The MLPA reactions were performed according to the manufacturer's general recommendations (http://www.mlpa.com). Briefly, 5 μl of genomic DNA (at a concentration of approximately 20 ng/μl) were incubated at 98°C for 5 min, cooled to room temperature and mixed with 1.5 μl of an appropriate probe mixture as well as 1.5 μl of SALSA hybridization buffer. The reaction was then denatured at 95°C for 2 min and hybridized at 60°C for 16 h. The hybridized probes were ligated at 54°C for 15 min by the addition of 32 μl of the ligation mixture. Following heat inactivation, the ligation reaction was cooled to room temperature, mixed with 10 μl of a PCR mixture (polymerase, dNTPs, and universal primers, one of which was labeled with fluorescein) and subjected to PCR amplification for 35 cycles.

The products of the MLPA reactions were diluted 20x in HiDi formamide containing GS Liz600, which was used as a DNA sizing standard, and separated via capillary electrophoresis (POP7 polymer) on an ABI Prism 3130XL apparatus (Applied Biosystems, Foster City, CA, USA). The obtained electropherograms were analyzed using GeneMarker software v2.4.0 (SoftGenetics, State College, PA, USA). The signal intensities (peak heights) were retrieved and transferred to prepared Excel sheets (available upon request). For each individual sample, the signal intensity of each probe was divided by the average signal intensity of the control probes to normalize the obtained values and to equalize the run-to-run variation. To calculate the relative signal (RS), the normalized signal of each probe was divided by the corresponding value of a reference (untreated) sample (or by the averaged value of the reference samples).

RESULTS

qEva-CRISPR strategy and assay design

For the proof-of-concept experiment, we designed two qEva-CRISPR assays (qEva_E4.1 and qEva_E4.2) specific to the INDEL scar-mutations generated by the well-validated E4.1 and E4.2 sgRNAs, which were developed previously (31). These sgRNAs target two non-overlapping sequences located in exon 4 of the *TP53* gene (Supplementary Table S2). Each of the assays included 8 control probes (specific to regions randomly distributed over the genome), one target-specific probe, and two (qEva_E4.1) or three (qEva_E4.2) off-target-specific probes. The off-target sequences, which were identified using Cas-OFFinder software (http://www.rgenome.net/cas-offinder/) (9), differed by 3–4 nucleotides from the corresponding target sequences (Supplementary Table S2). The set of control probes is universal and may be used in any qEva-CRISPR assay (except assays in which the control probes overlap with the tested target or off-target sites). The exact genomic location and sequence of each probe are indicated in Supplementary Table S3. The probes and general assay layouts were designed and generated according to a strategy developed for MLPA analysis that has been previously described in detail (27,28). We validated the performance of the assays with a panel of reference DNA samples and showed that each probe correctly recognized its target in the wild-type sequence [i.e. generates single PCR product (electropherogram peak) of the expected size].

qEva-CRISPR takes advantage of the fact that only perfectly hybridized MLPA half-probes can be ligated. Therefore, even a small mutation at the ligation point (predicted cut-site) prevents ligation and subsequent amplification of the probe. The amplification signal (amount of PCR product) is proportional to the dosage of matched (not-mutated) DNA molecules. Therefore, as shown in Figure 1A, the signal obtained for the target-specific probe in the DNA sample from cells treated with CRISPR/Cas9 reagents should be reduced proportionally to the fraction of mutated target sequences. In the case of modification of all targeted molecules, no signal of the target-specific probe should be present, whereas the signal for the control probes should remain intact (Figure 1A).

Using the developed assays, we analyzed HCT116 cells transfected with either the Cas9_E4.1 or Cas9_E4.2 plasmids. As shown in Figure 1B and Supplementary Figure S1, the relative signal (RS) of the target-specific probes decreased substantially in cells transfected with either of the plasmids. The level of the signal decrease corresponds to the fraction of modified target sequences expressed as the target modification efficiency (TME), which equals 49% (Supplementary Figure S1) and 43% (Figure 1B) for the E4.1 and E4.2 target sequences, respectively. In contrast, transfection with either of the plasmids did not affect the signals of off-target-specific probes (∼1, comparable to the signals of the control probes). This result indicated that with the conditions and sgRNAs that were used, the off-target sequences were not substantially edited. It must be noted; however, that some low-level (<5%) mutations in the tested regions may be missed. It results from the fact that to avoid false-positive detection of CRISPR/Cas9-induced mutations, signal changes smaller than 2 standard deviations (SDs) of the signals of control probes (usually smaller than 0.1 of RS or 10% of the TME value) were not considered significant.

Our experience, as well as the previous experiences of others, indicates that even the smallest mutation (mismatch of the probe with its target sequence) at the ligation point completely prevents ligation of the half-probes, which results in lack of signal for the affected probe [e.g. (27,37,38)]. To confirm this effect in our experimental system, we mod-

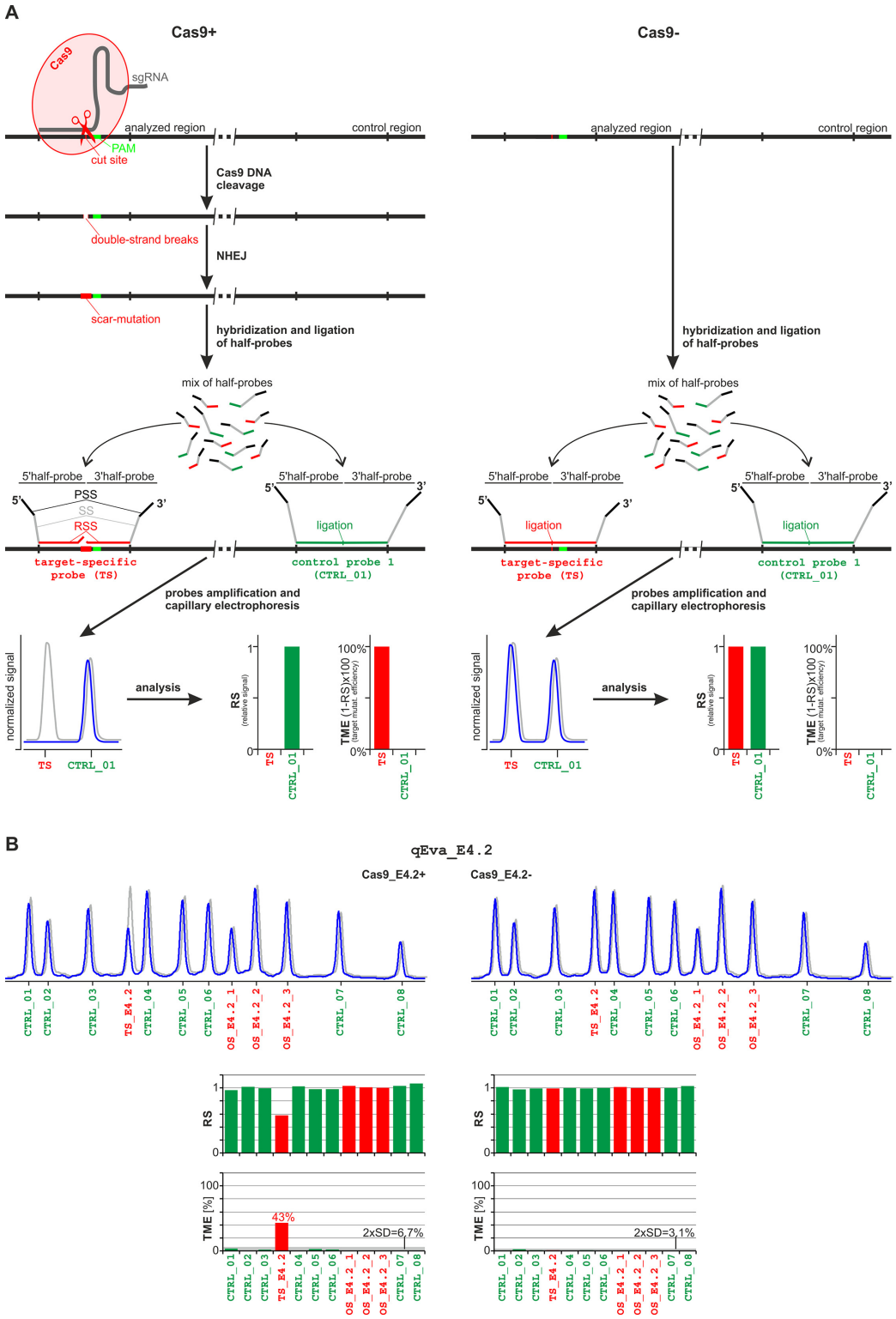


Figure 1. The qEva-CRISPR analysis strategy. (**A**) Schematic representation of qEva-CRISPR analysis of DNA samples extracted from cells either transfected (Cas9+; left-hand side) or not transfected (Cas9–; right-hand side) with a plasmid encoding CRISPR/Cas9 reagents. The Cas9 nuclease is directed toward target DNA by sgRNA, and after PAM recognition, DNA is specifically cleaved ∼3-nt upstream of PAM. Double-strand breaks (DSBs) are mainly repaired by NHEJ, resulting in INDEL scar-mutations. Target and off-target-specific qEva-CRISPR half-probes are designed to directly adjust the predicted cut-sites. Therefore, the occurrence of any scar-mutation prevents ligation of sister half-probes and subsequent probe amplification. Each half-probe is composed of a region-specific sequence [RSS, indicated by either red (TS, target-specific probe) or green (CTRL, control probe)]; stuffer sequence (SS, gray), and primer-specific sequences [PSS, specific to either forward (fluorescently labeled) or reverse universal primers)]. Note that different lengths of

ified the target-specific probe TS_E4.1 by introducing either a single substitution or single-nucleotide deletion into one of its half-probes (for details, see Supplementary Table S3 and Supplementary Figure S2). As shown in Supplementary Figure S2, any of the single-nucleotide mutations caused complete reduction of probe signal (i.e., no detectable signal in the expected position).

In addition, we designed an assay (qEva_SNP) specific for three different SNPs that served as a model for very small [single-nucleotide (sn) substitution, sn deletion, and sn insertion] CRISPR/Cas9-induced alterations. As shown in Supplementary Figure S3, qEva-CRISPR was able to specifically and quantitatively recognize all nucleotide alterations. The experiment confirmed that even small alteration at the ligation point prevents any signal of qEva-CRISPR probes. The experiment also confirmed the multiplexing capability of qEva-CRISPR. Further, to test the sensitivity and threshold of qEva-CRISPR mutation detection, we performed a titration experiment with decreasing amounts of sample containing the selected SNPs. As shown in Supplementary Figure S3B, the signal of a particular alteration still may be detected even if it is present in low fraction (∼2%). However, mutations detected at a very low level (<5%) should be interpreted carefully (which applies to most other methods as well).

Analysis of the well-characterized *VEGFA*, *CCR5* and *EMX1* sgRNAs activity with qEva-CRISPR

To demonstrate the utility and robustness of qEva-CRISPR for a wider range of targets, we designed five new assays that were specific for 5 different targets and corresponding off-targets, i.e. sgRNAs for *VEGFA* and *EMX1* and three sgRNAs for *CCR5*. The CRISPR/Cas9 activity towards the selected targets was analyzed before by T7E1 and NGS (33,39,40). As shown in Supplementary Figure S4A–C, the designed assay detected different levels of modification of target and off-target sites that generally corresponded to the results of previous studies (33,39,40). Please note that some of the targets (*CCR5*) are segmentally duplicated in other parts of the genome, which creates highly similar sequences (off-targets) (Supplementary Table S2). The high similarity of these sequences results in almost identical editing efficiency of the *CCR5* targets and the corresponding off-targets. A similar effect was also observed for the *EMX1* target but not for the *VEGFA* target, which differed more substantially from the off-target (compare Supplementary Figure S4 and Supplementary Table S2). To demonstrate the multiplexing ability of qEva-CRISPR, we designed an assay (qEva_multiplex) for parallel analysis of 3 targets (VEGFA, EMX1, and CCR5.7). We used qEva_multiplex for analysis of K562 cells, which were simultaneously transfected with three plasmids encoded the corresponding sgRNAs. (Supplementary Figure S4D).

Evaluation of experimental conditions that affect genome editing

To determine how the concentration of the plasmid encoding CRISPR/Cas9 components affects the efficiency of the target mutation, we transfected HCT116 and HeLa cells with increasing amounts (0.25–2 μg) of the Cas9_E4.1 and Cas9_E4.2 plasmids. As shown in Figure 2A, the fractions of the modified target sequences gradually increased from ∼0% at the lowest concentration to ∼35% at the highest concentration of plasmid. The efficiency of the target modification was comparable for the Cas9_E4.1 and Cas9_E4.2 plasmids in both HCT116 and HeLa cells. The signal of the off-target probes remained unchanged regardless of the concentration of plasmid and the type of transfected cells. The obtained qEva-CRISPR results corresponded well to the fraction of transfected cells observed using fluorescence microscopy (Figure 2A, insets) and correlated well (R^2 = 0.88–0.99) with the fraction of GFP-positive cells counted using flow cytometry (Figure 2B). It must be stressed, however, that the obtained results are specific to the experimental conditions and cannot be generalized to other sgRNAs or cell types.

In the next step, we analyzed the efficiency of the target modification in GFP-positive HCT116 cells fractionated by flow cytometry into three groups with high, medium, and low GFP fluorescence intensity. The experiment was performed with cells that were either transfected (Lipofectamine LTX) or electroporated with the Cas9_E4.2 plasmid. As shown in Figure 3, the efficiency of the target modification differed slightly depending on the method of transfection (i.e. higher for electroporation than lipofection). As expected, TME was generally higher in the fraction of GFP-positive cells than in unsorted cells (for comparison, see Figure 2; for cell-sorting parameters, see Supplementary Figure S5). The results are generally consistent with the efficiency of the target modification evaluated using the method based on T7E1 heteroduplex cleavage (Figure 3B).

Targeted gene modification in single-cell-derived clones

In the next step, we analyzed the level of target modification in single-cell-derived clones of HCT116 cells trans-

←

SSs allow for probes length differentiation and their separation by capillary electrophoresis. The signals of all of the probes were normalized (divided) to the average signal of the control probes (CTRL) and then to calculate the relative signal (RS) compared with the corresponding signals in a reference (untreated) sample. The signals of each probe are presented as bar plots, either as the RS or percent of the target mutation efficiency (TME = (1 – RS) × 100%). (**B**) Analysis of the target (*TP53*) and off-target mutation efficiencies in HCT116 cells either transfected (1.5 μg of plasmid DNA) or not transfected with the Cas9_E4.2 plasmid. From the top: overlapped electropherograms of the analyzed (blue) and reference (gray) samples, corresponding RS, and TME bar plots. The electropherograms were generated using the GeneMarker program (SoftGenetics, State College, PA) and captured as screenshots. The electropherogram lines were then extracted from the screenshot-bitmaps using the 'Trace line' tool in Corel Graphic v.X8. The probe IDs (TS – target-specific probe; OS – off-target-specific probes, CTRL – control probes) are indicated under the electropherograms and under the bar plots. The gray zone in the TME graphs indicates the significance threshold, 2× the standard deviation (SD) of the control probes signals. Target and off-target probes are indicated in red; control probes are indicated in green. The decrease in the signal of the TS probe corresponded to a TME of 43%, whereas the signal from two off-target specific probes remained unchanged (no mutation was detected). Corresponding results for the Cas9_E4.1 plasmid are shown in Supplementary Figure S1.

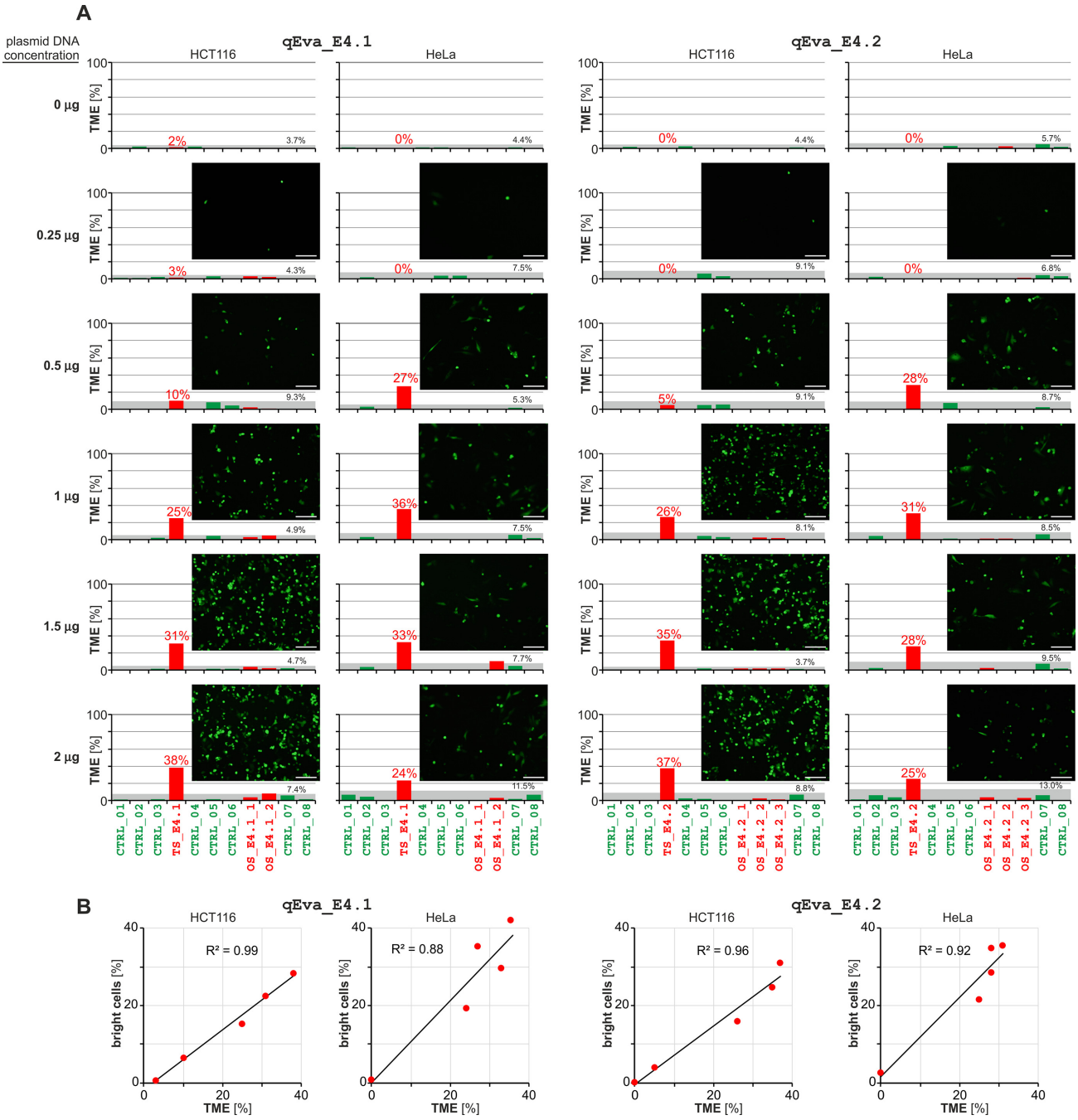


Figure 2. Effects of plasmid concentration on the efficiency of Cas9-induced target modification. (**A**) TME bar plots depicting the results of qEva-CRISPR analysis of HCT116 and HeLa cells transfected with increasing concentrations of either Cas9_E4.1 (left-hand side) or Cas9_E4.2 (right-hand side) plasmids. The chart designations are as shown in Figure 1B. The insets in each graph depict representative images of transfected cells (green fluorescent spots) expressing GFP as a marker. Scale bars, 50 µm. (**B**) Correlation of the TME values (x-axis) determined by qEva-CRISPR analysis (above) with the percentage of GFP-positive cells (y-axis) observed after plasmid transfection. GFP signals were detected in FL1 with the standard BD Accuri C6 filter configuration FL1 = 533/30 nm. Gates were set based on the untransfected controls.

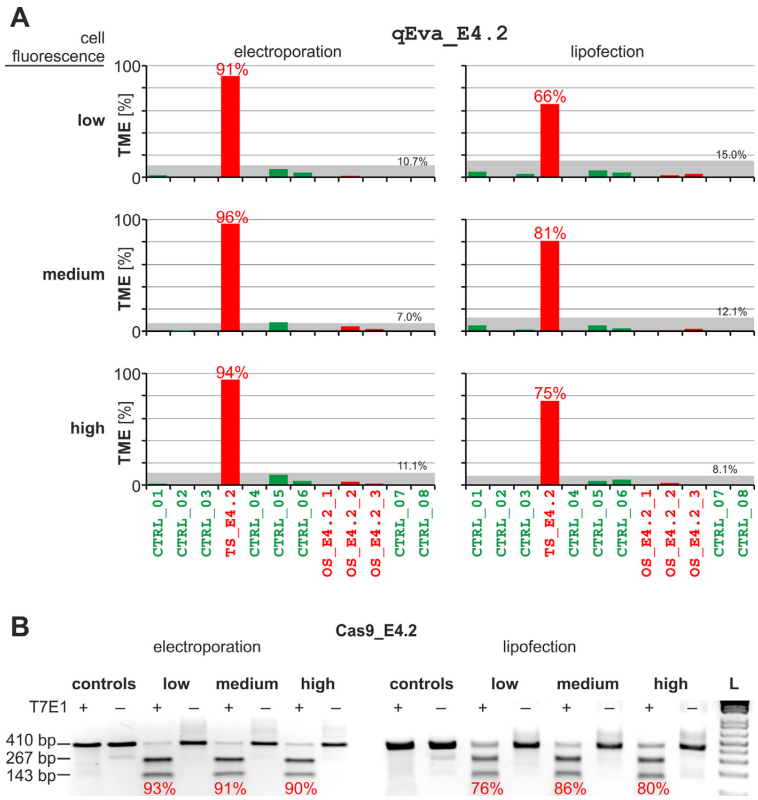


Figure 3. Analysis of the Cas9-induced target modification efficiency in GFP-positive cells. The GFP signal indicates cells carrying the Cas9_E4.2 plasmid transfected either by electroporation (left-hand side) or lipofection (right-hand side). Based on the fluorescence intensity, GFP-positive cells were divided into three arbitrarily designated categories: low, medium, and high. (**A**) The TME bar plots depicting the results of qEva-CRISPR analysis of low, medium, and high fractions of the HCT116 GFP-positive cells. The chart designations are as shown in Figure 1B. (**B**) Agarose gels depicting the results of T7 endonuclease I (T7E1) analysis of DNA samples from the above-described cells. Bands representing full-length (410 bp) and digested PCR products (∼267 bp and 143 bp) are indicated on the left. Only heteroduplexes with Cas9-induced mismatches are digested by T7E1. PCR products digested (+) and not digested (–) with T7E1 are shown next to each other. Control – PCR product of DNA samples extracted from cells not transfected with Cas9_E4.2 plasmid; L – 1 Kb Plus DNA ladder (ThermoFisher Scientific, Waltham, MA). The percentage of digested products (calculated based on the densitometric analysis) corresponding to a fraction of the Cas9-mutated targets are shown in the particular lines on agarose gels (red fonts).

fected with either Cas9_E4.1 or Cas9_E4.2 plasmids. As shown in Figure 4A, the level of target modification in both cases was either close to 0% or close to 100%, which suggested that either no allele was mutated or both alleles of the analyzed cells were mutated. There were no intermediate values (close to 50%) that would indicate mutation of just one allele (heterozygous mutation). The results of the Sanger sequencing analysis of selected clones were perfectly consistent with the observed genotypes (which confirm either wild-type or biallelic mutations) (Figure 4B). In contrast, the levels of off-target mutations were consistently close to 0%, which indicates there was no mutation of the analyzed off-targets in the tested clones. It must be noted, however, that in some clones with biallelic mutations, the signal of the target-specific probes was not completely reduced (RS > 0), and it corresponds to TME values close but not equal to 100% (Figure 4A). The occurrence of some signal of target-specific probes in single-cell-derived clones with biallelic mutations may result either from some contamination of the analyzed clone with wild-type cells (clone cross-contamination) or from the occurrence of a mutation that allows for some low-level hybridization and subsequent ligation of the probe. The second explanation seems to be less likely (see Supplementary Figure S2).

Application of qEva-CRISPR for analysis of 'difficult' genomic regions

We tested the applicability of qEva-CRISPR for analysis of 'difficult' regions enriched in low complexity and highly repetitive sequences. Such regions are generally very difficult or not possible to analyze using standard genetic approaches. For this purpose, we designed three qEva-CRISPR assays (i.e. qEva_HTT_sg1, qEva_HTT_sg3, and qEva_HTT_sg4) to analyze the targets located in exon 1 of the *HTT* gene in close proximity to polymorphic microsatellite tracts (CAG and CCG) (Figure 5A and B), the expansion of one of which (CAG) causes Huntington's disease. These assays utilized a set of previously designed control probes. As shown in Figure 5B, some half-probes of the target-specific probes showed substantial overlap with highly redundant repetitive sequences. The specificity of these probes was guaranteed by combination with a second, more specific half-probe. The experiment performed in four biological replicates, clearly showed that each of the applied

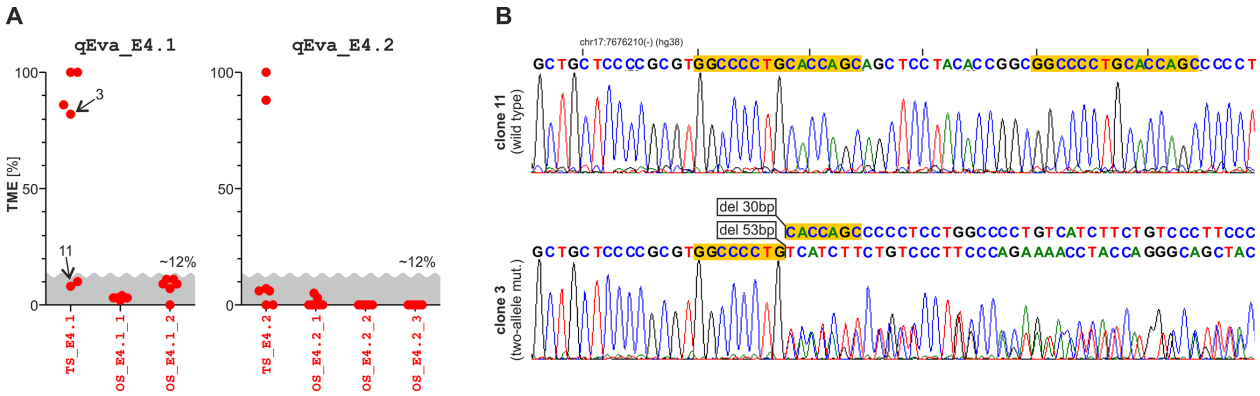


Figure 4. Analysis of the level of target modifications in single-cell-derived clones of HCT116 cells transfected with either Cas9_E4.1 or Cas9_E4.2 plasmids. (**A**) A dot-plot depicting the TME values of target- and off-target-specific probes. Each dot represents the TME value of one analyzed clone. The gray zone indicates the approximate level of 2x the SD observed in the analyzed clone samples. (**B**) Sanger sequencing analysis of the Cas9_E4.1 target site in clone 11 with the wild-type sequence and clone 3 with biallelic deletions in the target sites. The dots representing clone 3 and clone 11 are indicated in the dot-plot presented in panel A. Yellow backgrounds indicate micro-homologous sequences flanking the Cas9_E4.1 target site.

sgRNAs efficiently induced mutations in the corresponding target with HTT.sg4 showing the highest efficiency (Figure 5C and D). The tested off-targets of the HTT.sg3 and HTT.sg4 sgRNAs were not affected. In contrast, both of the tested off-targets (HTT.sg1_1 and HTT.sg1_2) of the HTT.sg1 sgRNA were substantially modified (Figure 5C and D). The high mutation rate of these off-targets was consistent with their high similarity to the target (single mismatch at the second nucleotide upstream of the expected cut-site) (Supplementary Table S2). For comparison to biological variation (Figure 5D), we performed technical replication (7×) of the selected experiment as well. As shown in our results (compare Figure 5D and E), the technical variation is very small, and therefore, most qEva-CRISPR variability results from biological components. Additionally, we replicated analysis of HTT.sg1 efficiency with the use of a ribonucleoprotein (RNP) complex delivery system. This dose-dependent experiment was performed in HEK293 cells with four biological replicates (Figure 5F). As shown in our results, the RNP delivery system generally resulted in lower target modification efficiency, but it also resulted in substantially lower modification of the off-targets.

Distinguishing HDR from NHEJ

Finally, we designed the qEva_HTT.edit assay to demonstrate that qEva-CRISPR can be applied not only to analyze the level of 'destruction' of specific sequences (loss of signal by Cas9 cutting) but also to detect and quantify new sequences (gain of signal) generated by NHEJ or HDR. In the first step, we used previously described HTT.sg1 and HTT.sg4 for precise excision of the DNA sequence between cuts generated by either Cas9 or Cas9n (34). As shown in Figure 6A–B, excision of the DNA fragment resulted in a decrease in the signal of the HTT.sg1 and HTT.sg4-specific probes. Simultaneously, the signal of the TS_HTT.sg1 /4 probe appears. The results indicate clean rejoining of the DNA free-ends by NHEJ. In addition, to demonstrate detection of new sequences introduced by HDR, we transfected HEK293T cells with Cas9(n)_HTT.sg1 and Cas9(n)_HTT.sg4 plasmids along with the donor template (ssODN containing fragment of GFP sequence). As shown in Figure 6C, qEva-CRISPR can detect and distinguish new sequences created either by NHEJ (with clean ends rejoining) or HDR. Analysis of single-cell-derived clones confirmed the ability of qEva-CRISPR to detect specific sequences resulting from various repair mechanisms (Figure 6D and E).

DISCUSSION

New CRISPR technologies for targeted genome editing are still being developed, which will enhance the range of applications and possible targets. A simple method for reliable and quantitative determination of INDEL mutations in a target sequence is important for the selection of effective approaches and CRISPR reagents, including targeting sgRNAs. Another important factor that requires careful evaluation is the specificity of the approach that is used, which is especially important for therapeutic applications and preclinical studies.

The typical result of the genome editing procedure is a heterogeneous mixture of different types of mutations occurring at a target site that is additionally accompanied by an unmodified wild-type target sequence. This heterogeneity substantially hampers reliable evaluation of editing efficiency, especially for quantification. The methods that allow for detection and quantification of CRISPR-induced mutations include NGS [e.g., (39)]; Tracking of Indels by DEcomposition (TIDE) [(21,22)]; Indel Detection by Amplicon Analysis (IDAA) (41); and the Droplet Digital PCR (ddPCR)-based method (42).

The advantage of NGS, especially whole-genome NGS, is its ability to detect the CRISPR-induced mutation in any region of the genome, not only mutations occurring in a predefined target or off-targets but also accidental mutations in regions that are difficult or impossible to predict. This characteristic makes NGS a method of choice for the final verification of editing outcomes, especially when it is used in the context of clinical applications. On the other hand, NGS is still costly, cumbersome and time-consuming (NGS library

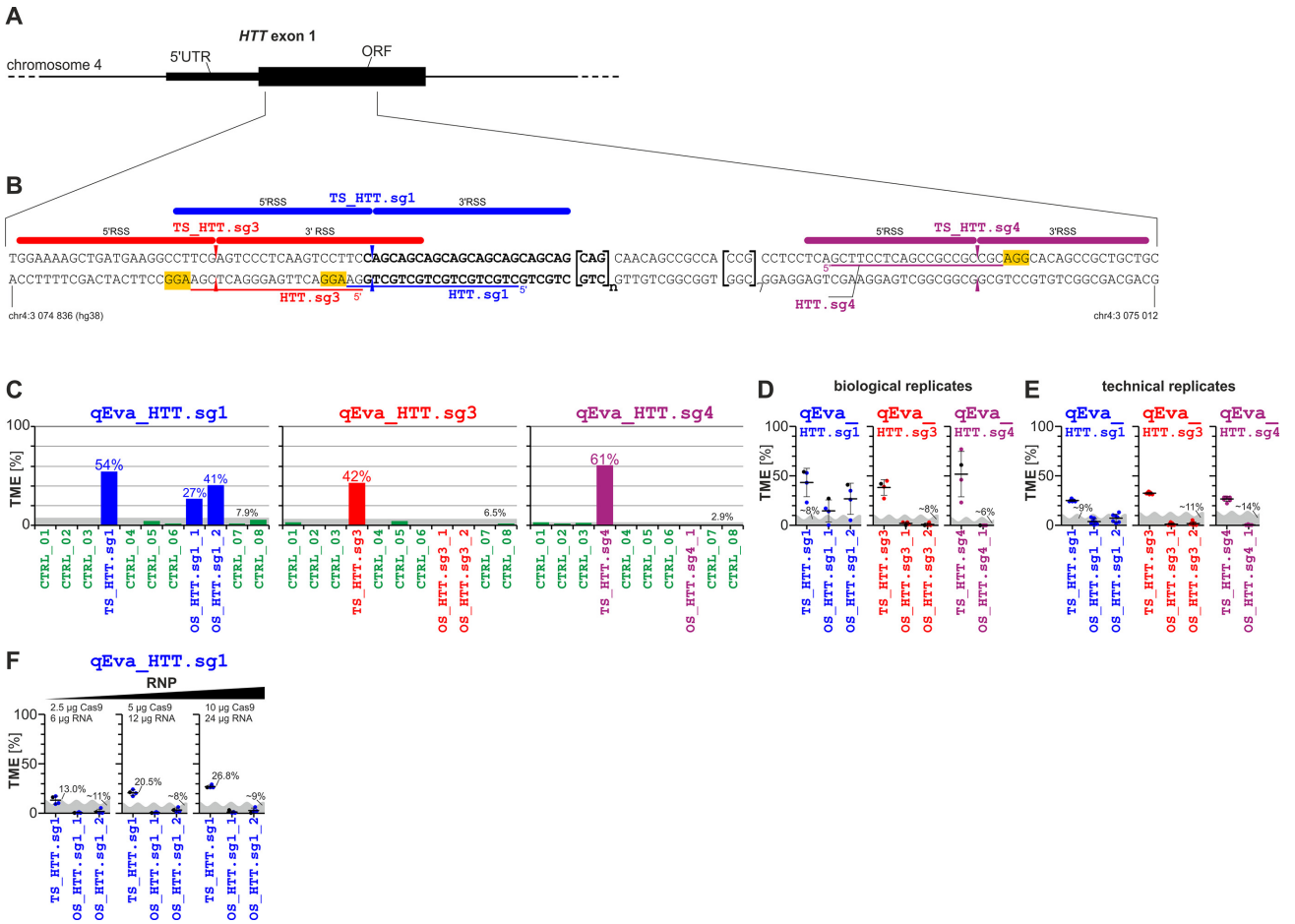


Figure 5. Assay design and analysis of the Cas9-induced target modification efficiency in the region of the *HTT* gene. (**A**) Schematic representation of exon 1 of the *HTT* gene with lower and higher rectangles representing the 5′UTR and coding sequence, respectively. (**B**) The sequence of *HTT* exon 1 with the indicated target sequences (thin lines below the targeted strand), corresponding to predicted cut-sites (arrowheads) and PAMs (yellow background). The 5′ and 3′ RSSs of the qEva-CRISPR probes specific to particular target sites are depicted above the sequence. Bold font indicates a CAG repeat sequence, which can lead to Huntington's disease if repeats are expanded. Please also note the presence of other repetitive and low-complexity elements in the sequence (e.g. CCG repeats). (**C**) The TME bar graphs depicts the level of the target and corresponding off-target modifications in the transfected HEK293T cells. Blue, red, and purple colors indicate the HTT.sg1, HTT.sg3 and HTT.sg4 target sequences as well as the corresponding qEva-CRISPR probes and target/off-target-specific bars in the TME graph, respectively. (**D**) Dot-plots summarizing results (TME values in the tested target/off-target sites) of four independent CRISPR/Cas9 experiments (biological replicates). Dots represent the TME values of individual experiments (black dots indicate the results shown in (C), horizontal lines represent the medians, and whiskers represent the SD values. (**E**) Dot plots summarizing the results (TME values in the tested target/off-target sites) of seven qEva-CRISPR analyses (technical replicates) for the same CRISPR/Cas9 experiment. The graphs scheme as shown in (D). (**F**) The results of four independent experiments of CRISPR/Cas9-induced sequence modifications performed with the use of HTT.sg1 delivered in increasing amounts (indicated over the graphs) of RNP complex. The dot plots scheme is the same as shown in (D), blue and black dots represent experiments performed in the HEK293T and GM04281 cells, respectively.

preparation and data analysis), and it is often performed with the use of an external service. Therefore, NGS is impractical and rarely used in everyday laboratory work. The methods that may overcome this gap are the TIDE, IDAA and ddPCR-based methods as well as qEva-CRISPR. All of these methods have a similar cost (∼$10 per sample; depending on the analysis setup) and require a similar amount of time for analysis (1–2 days). All of them need only basic and easily accessible laboratory equipment, such as a thermocycler and either a CE-apparatus (TIDE, IDAA, qEva-CRISPR) or ddPCR equipment dedicated to droplet generation and counting. The TIDE method is based on the decomposition of Sanger sequencing traces to identify major mutations induced in the expected editing sites. The main advantage of TIDE is that it simultaneously identifies and quantifies all major mutations occurring at the target site. It was shown that TIDE results excellently correlate (predict) with the global effects of target modification. However, the estimated level of individual mutations may deviate from expected (from replicate or control experiments) by a few percentage points, and it is somehow less reliable for mutations occurring at a low frequency as well as for larger INDELs. Inherently, the reliability of TIDE depends on the purity/quality of PCR products and Sanger sequencing reads, and therefore, sequencing the opposite strand to confirm the results is recommended. IDAA is a very simple and straightforward method that takes advantage of the high resolution (1 nt) of capillary electrophoresis for sepa-

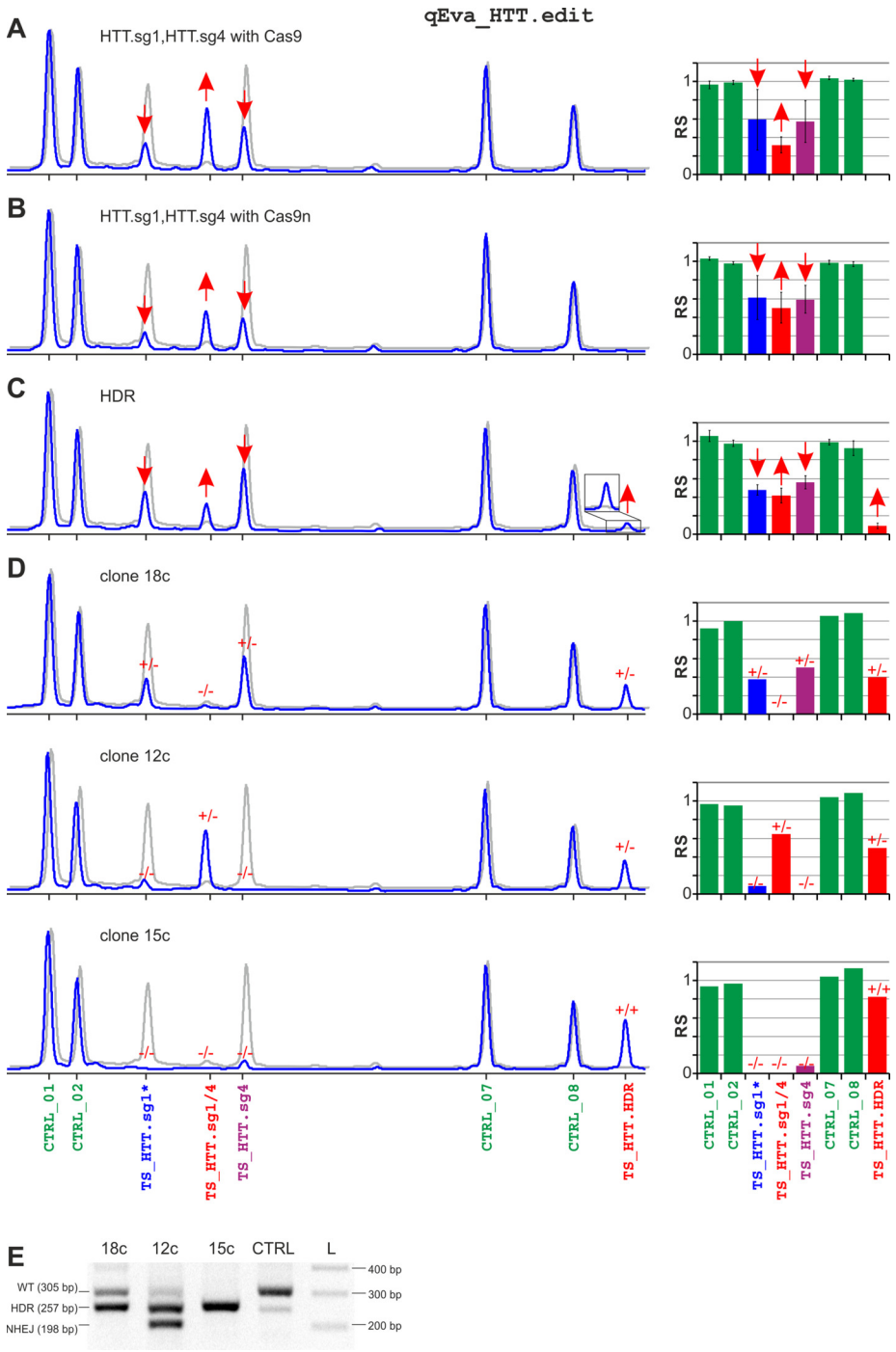


Figure 6. Analysis of NHEJ and HDR events after targeted editing of the *HTT* gene by CRISPR/Cas9. (**A**) Results of the HTT gene editing with the use of two sgRNAs (HTT.sg1 and HTT.sg4, as shown in Figure 5) flanking the (CAG)n repeat tract; (left-hand-side) electropherograms of the qEva_HTTedit assay of treated (blue) versus untreated (reference; gray) samples; (right-hand-side) RS-bar plot summarizing the results of 3 independent experiments (error bars indicate SD). Note that the excision of a genomic fragment between the HTT.sg1 and HTT.sg4 cuts destroys both targets and decreases the signal of both target-specific probes. Subsequent rejoining of the free-ends by NHEJ creates a new sequence that is recognized by the TS_HTT.sg1/4 probe that is composed of the 5′ half probe of TS_HTT.sg1* and the 3′half probe of TS_HTT.sg4 (see Supplementary Table S3). The signal of TS_HTT.sg1/4 is not present in a reference sample and increases after the treatment. (**B**) Similar to (A) but conducted with Cas9n (instead of Cas9). (**C**) Similar to (A) but conducted with an ssODN composed of a 59-nt insert sequence and 60-nt 5′ and 3′ arms homologous to the HTT.sg1 and HTT.sg4 flanks. The appearance of the HTT_HDR probe signal indicates the successful introduction of the insert by HDR. (**D**) Examples of single-cell-derived clones generated from the experiment as shown in (C): (from the top) the clone with one allele with wild-type sequence and one allele modified by HDR; one allele modified by HDR and the other by NHEJ; and the clone with two alleles modified by HDR. (E) The agarose gel analysis of the clones shown in (C) (clone IDs are shown above the gel). The PCR product was generated with primers flanking the modification site. WT – the PCR product of wild-type sequence; HDR – the product of sequence repaired by HDR; and NHEJ – the product of sequence repaired by NHEJ (with clean ends rejoining).

ration of all INDEL mutations. Its limitation, however, is that it cannot detect substitutions, which means that it (in contrast to TIDE and qEva-CRISPR) cannot be used for analysis of the efficiency of recently developed strategies of targeted base editing [e.g. (43) and references within]. Similar to TIDE and other PCR-based methods (unlike qEva-CRISPR) IDAA does not allow for detection of larger rearrangements extending beyond amplicon size. Additionally, IDAA relies on the analysis of a homogenous PCR product. Therefore (like TIDE), it is less suitable for regions with overlapping tandem repeats and INDEL polymorphisms. A recent direct comparison of T7E1, TIDE and IDAA with NGS used as a gold standard showed that both TIDE and IDAA substantially outperform T7E1 that turned out to poorly predict editing efficiency (44). The results also showed that both TIDE and IDAA overestimate (10–20%) the presence of wild-type sequences and miscall some mutations, especially 'complex' ones. Despite their specific limitations, both TIDE and IDAA are presently methods of choice for standard analysis of genome editing. However, both methods are less suitable for analysis of targets located in 'difficult regions' and they have little potential for multiplexing.

More recently, two strategies (based on similar principles) of genome editing evaluation that take advantage of the partitioned amplification of individual copies of DNA with the use of ddPCR have been proposed (42,45). The ddPCR-based method utilizes two fluorescently labeled probes that are specific for closely located sequences (one close to the wild-type reference site and the other close to the expected target-cut-site). The absolute quantification of PCR droplets positive for one and two probes allows for evaluation of the fraction of edited DNA molecules. The ddPCR-based method potentially allows for detection of a very small fraction of mutated molecules. However, the performance of this method strongly depends on assay conditions and the amount of DNA used for analysis (both factors require optimization and may affect the precision of analysis) (Findlay SD). It may be expected that improvement of ddPCR technology (both in terms of a number of analyzed droplets and assay conditions) may further enhance the sensitivity and reliability of the ddPCR-based method and make it a method of choice to detect mutations present in a very small fraction of analyzed molecules.

The main advantages of qEva-CRISPR compared to these methods include multiplexing, its relative insensitivity to mutation type, the capability for analyzing targets located in 'difficult regions' and for distinguishing HDR and NHEJ. We tested the qEva-CRISPR strategy and confirmed its reliability using ten CRISPR/Cas9 targets located in the *TP53, VEGFA*, *EMX1* and *CCR5* genes as well as the highly repetitive region of the *HTT* gene. Reliable analysis of Cas9-mediated mutations in the latter region was not possible using other PCR-based or sequencing-based methods. The designed qEva-CRISPR assays were tested in four different cell types under numerous experimental conditions and designs (i.e. different plasmid and RNP concentrations, transfection methods, sorted and unsorted cells, and single-cell-derived clones) as well as artificial models. To test the reliability of qEva-CRISPR, we compared the obtained results with (i) the number of positively transfected (GFP-positive) cells observed either with fluorescence microscopy or counted with a flow cytometer, (ii) the results obtained for T7E1 analysis, (iii) Sanger sequencing of the target sequences in single-cell-derived clones and (iv) the results of previous studies. All of the performed analyses showed substantial consistency with the qEva-CRISPR results.

The target-specific qEva-CRISPR probes were designed to overlap target sequences with the ligation position located directly over the predicted cut-sites (Figure 1A, also B), which makes them (and qEva-CRISPR analysis) sensitive to any type of mutation occurring at the target sites, including both small mutations (even single-nucleotide substitutions or deletions) and large deletions or more complex rearrangements. It has been previously shown [e.g. (27,37,38)] and confirmed herein (Supplementary Figure S2 and Supplementary Figure S3) that even a small mismatch at the ligation point completely prevents ligation and subsequent amplification of the probe. It was also shown that the signal of the MLPA probes is quantitatively proportional to the dosage (number of copies) of the tested regions [e.g. (46–48)]. Sensitivity to large deletions is the main property of the ligation-dependent probes that are routinely utilized in MLPA assays for large mutation detection [e.g. (27,29,49,50)]. Moreover, the short region-specific sequence (usually ∼40 nt) of the qEva-CRISPR probes utilized in our study made them relatively insensitive to the single-nucleotide substitutions (SNPs) that commonly occur in the human genome as well as in other genomes. Because the most common SNPs in the human genome have already been identified and deposited in the appropriate databases (e.g. dbSNP), qEva-CRISPR probes may, in most cases, be designed to avoid common SNPs in a relatively short region-specific sequence [see protocol in (28)]. The occurrence of SNPs is a serious drawback of all heteroduplex-based methods, which usually use relatively long PCR amplicons (∼400 nt) that often overlap with common SNPs and interfere with subsequent analyses.

The experiments that were performed also confirmed that qEva-CRISPR analysis may be used for optimization and selection of the best experimental conditions (plasmid and RNP concentration or transfection method) for most effective target modifications or for the selection of single-cell-derived clones with the desired genotype, including specific genotypes generated by HDR or NHEJ without the need for cumbersome and time-consuming cloning and sequencing of individually targeted alleles.

The advantages of our qEva-CRISPR assays are as follows. (i) qEva-CRISPR is a multiplex method that enables simultaneous analysis of target and off-target regions. In the same way, it could be adapted for simultaneous analysis of more than one target site. (ii) qEva-CRISPR is a semi-quantitative method that allows for robust estimation of the fraction of modified targets. (iii) qEva-CRISPR may be designed to detect and distinguish new sequences generated by NHEJ and HDR. (iv) The sensitivity of qEva-CRISPR does not depend on the mutation type. The qEva-CRISPR probe may detect both single-nucleotide substitutions and deletion of the whole gene. (v) qEva-CRISPR allows for detection of homozygous mutations without the need for the addition of wild-type DNA, which is required in all methods that rely on heteroduplex formation. (vi) qEva-

CRISPR does not require the optimization and adjustment of assays for the specific target sequence. It takes advantage of a standard protocol (standard reaction conditions, easily accessible reagent set) of MLPA. The standard MLPA setup was validated and successfully used in hundreds of research and clinical studies for the analysis of large mutations in disease-related genes [e.g. (27,29,49,50)]. (vii) The qEva-CRISPR strategy can be easily adapted to the target or off-targets of interest. This approach requires only the design of target-specific probes that may be used together with a set of control probes designed for this study (upon request, we may share aliquots of a ready-to-use mix of control probes). (viii) The qEva-CRISPR test is cost-effective (∼$10 per sample, including the reagents and cost of the capillary electrophoresis separation (excluding the starting cost of probes synthesis, which is ∼$150 per probe, the synthesized probe may be used for thousands of reactions). (ix) qEva-CRISPR does not require advanced laboratory equipment. (x) Finally, due to the characteristics of the target-specific probes and the strategy utilized for the probe design [described in detail previously by our group (27,28)], qEva-CRISPR may be used to analyze almost any target or targets of interest, including targets located in 'difficult' genomic regions.

It must also be noted that qEva-CRISPR has limitations. It cannot be used for a whole-genome analysis and does not allow for detection of all potential and accidental off-target sites, and it generally does not characterize the identified mutations. qEva-CRISPR does not allow for reliable detection of a mutation that occurs in a very small fraction (∼5%) of targets. However, because in most experiments, only results in which the target of interest is substantially modified are considered important and worthy of further analysis, the latter drawback has a limited impact on this method.

In summary, the plethora of methods developed for evaluating genome editing efficiency (including methods not discussed here) indicates there is no ideal method that fulfills all requirements for testing the efficiency. The methods discussed in this study have different advantages and are optimal for different applications. We may expect that in the future, either one method will turn out to be superior and will dominate the others or, more likely, different methods will form niches for specific applications.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

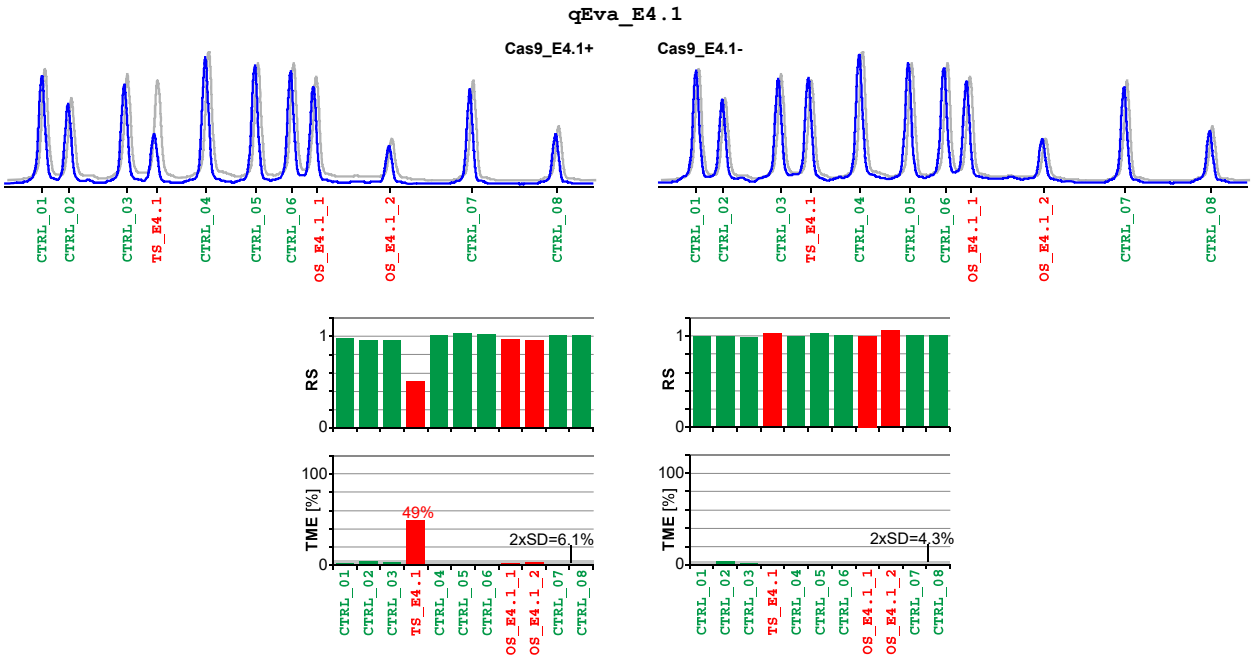
FUNDING

National Science Center PL [2015/18/E/NZ2/00678 and 2016/22/A/NZ2/00184]; Ministry of Science and Higher Education PL [KNOW program for years 2014–2018]. Funding for open access charge: National Science Center [2015/18/E/NZ2/00678].
Conflict of interest statement. None declared.

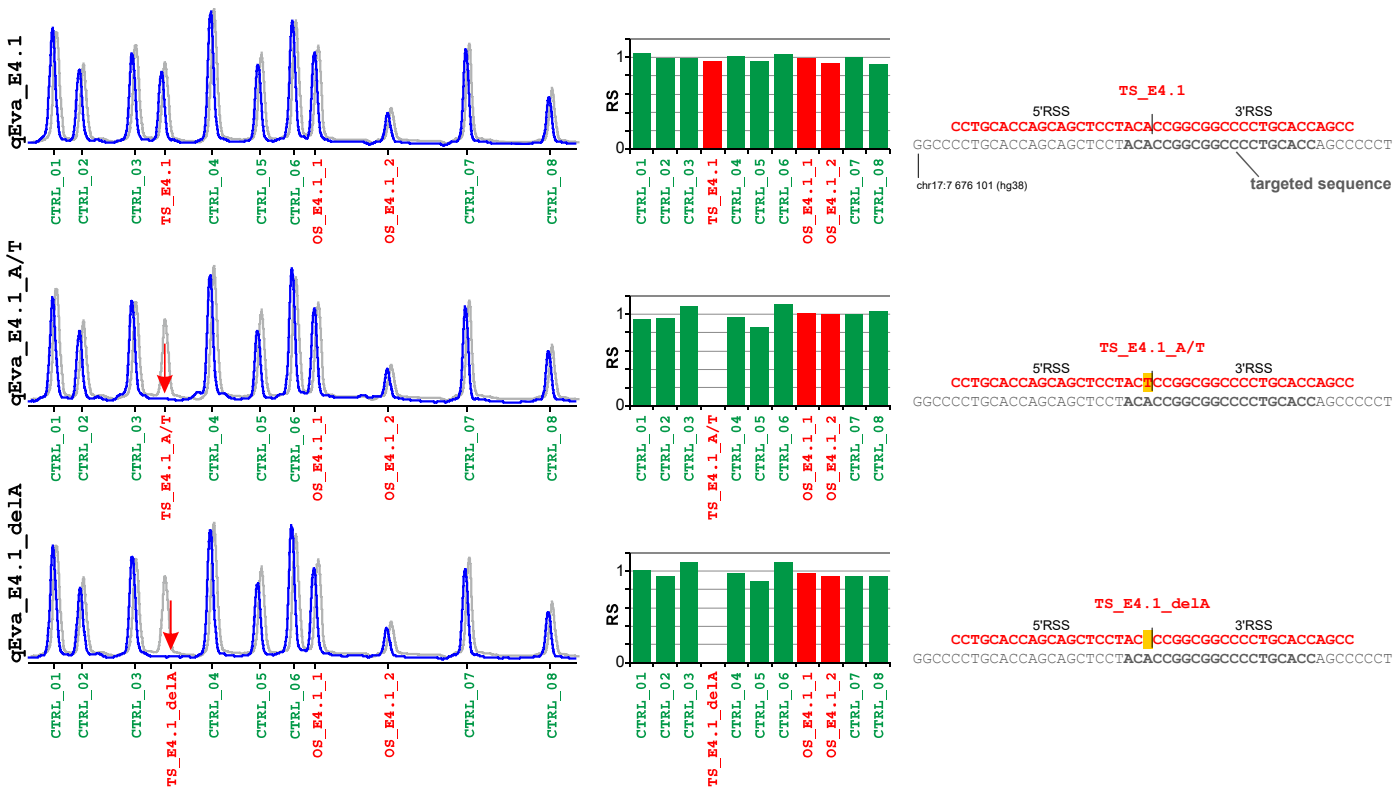
REFERENCES

1. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A Programmable Dual-RNA-Guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
2. Ran,F.A., Hsu,P.D., Wright,J., Agarwala,V., Scott,D.A. and Zhang,F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–2308.
3. Mojica,F.J.M., Diez-Villasenor,C., Garcia-Martinez,J. and Almendros,C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, **155**, 733–740.
4. Anders,C., Niewoehner,O., Duerst,A. and Jinek,M. (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, **513**, 569–573.
5. 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. and Holmes,M.C. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, **435**, 646–651.
6. Dahlem,T.J., Hoshijima,K., Jurynec,M.J., Gunther,D., Starker,C.G., Locke,A.S., Weis,A.M., Voytas,D.F. and Grunwald,D.J. (2012) Simple methods for generating and detecting Locus-Specific mutations induced with TALENs in the zebrafish genome. *PLoS Genet.*, **8**, e1002861.
7. Jiang,F.G. and Doudna,J.A. (2017) CRISPR-Cas9 structures and mechanisms. *Annu. Rev. Biophys.*, **46**, 505–529.
8. Moreno-Mateos,M.A., Vejnar,C.E., Beaudoin,J.D., Fernandez,J.P., Mis,E.K., Khokha,M.K. and Giraldez,A.J. (2015) CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods*, **12**, 982–988.
9. Bae,S., Park,J. and 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.
10. Erard,N., Knott,S.R.V. and Hannon,G.J. (2017) A CRISPR resource for individual, combinatorial, or multiplexed gene knockout. *Mol. Cell*, **67**, 348–354.
11. Tsai,S.Q. and Joung,J.K. (2016) Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat. Rev. Genet.*, **17**, 300–312.
12. Zischewski,J., Fischer,R. and Bortesi,L. (2017) Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnol. Adv.*, **35**, 95–104.
13. Kim,Y., Kweon,J., Kim,A., Chon,J.K., Yoo,J.Y., Kim,H.J., Kim,S., Lee,C., Jeong,E., Chung,E. *et al.* (2013) A library of TAL effector nucleases spanning the human genome. *Nat. Biotechnol.*, **31**, 251–258.
14. Qiu,P., Shandilya,H., D'Alessio,J.M., O'Connor,K., Durocher,J. and Gerard,G.F. (2004) Mutation detection using Surveyor nuclease. *Biotechniques*, **36**, 702–707.
15. Zhu,X.X., Xu,Y.J., Yu,S.S., Lu,L., Ding,M.Q., Cheng,J., Song,G.X., Gao,X., Yao,L.M., Fan,D.D. *et al.* (2014) An efficient genotyping method for Genome-modified animals and human cells generated with CRISPR/Cas9 system. *Sci. Rep.*, **4**, 6420.
16. Vouillot,L., Thelie,A. and Pollet,N. (2015) Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3 (Bethesda)*, **5**, 407–415.
17. Kim,H., Um,E., Cho,S.R., Jung,C., Kim,H. and Kim,J.S. (2011) Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat. Methods*, **8**, 941–943.
18. Kim,J.M., Kim,D., Kim,S. and Kim,J.S. (2014) Genotyping with CRISPR-Cas-derived RNA-guided endonucleases. *Nat. Commun.*, **5**, 3157.
19. Yu,C., Zhang,Y.G., Yao,S.H. and Wei,Y.Q. (2014) A PCR Based protocol for detecting indel mutations induced by TALENs and CRISPR/Cas9 in Zebrafish. *PLoS One*, **9**, e98282.
20. Bauer,D.E., Canver,M.C. and Orkin,S.H. (2015) Generation of genomic deletions in mammalian cell lines via CRISPR/Cas9. *J. Vis. Exp.*, **83**, e52118.
21. Brinkman,E.K., Chen,T., Amendola,M. and van Steensel,B. (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.*, **42**, e168.
22. Brinkman,E.K., Kousholt,A.N., Harmsen,T., Leemans,C., Chen,T., Jonkers,J. and van Steensel,B. (2018) Easy quantification of template-directed CRISPR/Cas9 editing. *Nucleic Acids Res.*, **46**, e58.
23. Guell,M., Yang,L.H. and Church,G.M. (2014) Genome editing assessment using CRISPR Genome Analyzer (CRISPR-GA). *Bioinformatics*, **30**, 2968–2970.
24. Carrington,B., Varshney,G.K., Burgess,S.M. and Sood,R. (2015) CRISPR-STAT: an easy and reliable PCR-based method to evaluate target-specific sgRNA activity. *Nucleic Acids Res.*, **43**, e157.

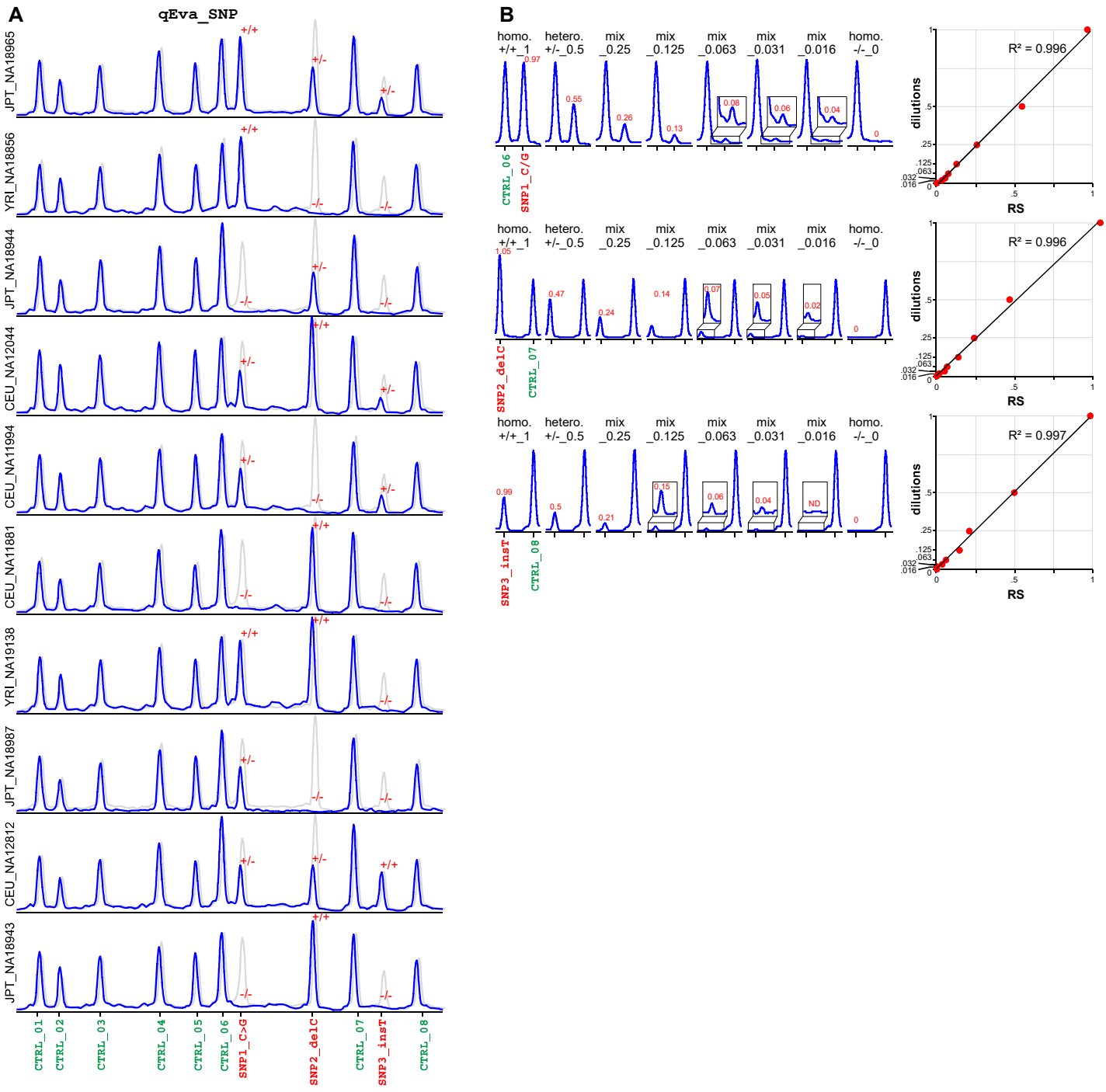
25. Kc,R., Srivastava,A., Wilkowski,J.M., Richter,C.E., Shavit,J.A., Burke,D.T. and Bielas,S.L. (2016) Detection of nucleotide-specific CRISPR/Cas9 modified alleles using multiplex ligation detection. *Sci.Rep.*, **6**, 32048.
26. Tycko,J., Myer,V.E. and Hsu,P.D. (2016) Methods for optimizing CRISPR-Cas9 genome editing specificity. *Mol. Cell*, **63**, 355–370.
27. Kozlowski,P., Roberts,P., Dabora,S., Franz,D., Bissler,J., Northrup,H., Au,K.S., Lazarus,R., Domanska-Pakiela,D., Kotulska,K. *et al.* (2007) Identification of 54 large deletions/duplications in TSC1 and TSC2 using MLPA, and genotype-phenotype correlations. *Hum. Genet.*, **121**, 389–400.
28. Marcinkowska,M., Wong,K.K., Kwiatkowski,D.J. and Kozlowski,P. (2010) Design and generation of MLPA probe sets for combined copy number and small-mutation analysis of human genes: EGFR as an example. *Scientific World J.*, **10**, 2003–2018.
29. Schouten,J.P., McElgunn,C.J., Waaijer,R., Zwijnenburg,D., Diepvens,F. and Pals,G. (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.*, **30**, e57.
30. Kozlowski,P., Jasinska,A.J. and Kwiatkowski,D.J. (2008) New applications and developments in the use of multiplex ligation-dependent probe amplification. *Electrophoresis*, **29**, 4627–4636.
31. Ramlee,M.K., Yan,T.D., Cheung,A.M.S., Chuah,C.T.H. and Li,S. (2015) High-throughput genotyping of CRISPR/Cas9-mediated mutants using fluorescent PCR-capillary gel electrophoresis. *Sci. Rep.*, **5**, 15587.
32. Cradick,T.J., Fine,E.J., Antico,C.J. and Bao,G. (2013) CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.*, **41**, 9584–9592.
33. Cho,S.W., Kim,S., Kim,Y., Kweon,J., Kim,H.S., Bae,S. and Kim,J.S. (2014) Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.*, **24**, 132–141.
34. Dabrowska,M., Juzwa,W., Krzyzosiak,W.J. and Olejniczak,M. (2018) Precise excision of the CAG tract from the huntingtin gene by Cas9 nickases. *Front. Neurosci.*, **12**, 75.
35. Serizawa,R.R., Ralfkiaer,U., Dahl,C., Lam,G.W., Hansen,A.B., Steven,K., Horn,T. and Guldberg,P. (2010) Custom-Designed MLPA using multiple short synthetic probes application to methylation analysis of five promoter CpG islands in tumor and urine specimens from patients with bladder cancer. *J. Mol. Diagn.*, **12**, 402–408.
36. Langerak,P., Nygren,A.O.H., Schouten,J.P. and Jacobs,H. (2005) Rapid and quantitative detection of homologous and non-homologous recombination events using three oligonucleotide MLPA. *Nucleic Acids Res.*, **33**, e188.
37. Lewandowska,M.A., Czubak,K., Klonowska,K., Jozwicki,W., Kowalewski,J. and Kozlowski,P. (2015) The use of a Two-Tiered testing strategy for the simultaneous detection of small EGFR mutations and EGFR amplification in lung cancer. *PLoS One*, **10**, e0117983.
38. Marcinkowska-Swojak,M., Handschuh,L., Wojciechowski,P., Goralski,M., Tomaszewski,K., Kazmierczak,M., Lewandowski,K., Komarnicki,M., Blazewicz,J., Figlerowicz,M. *et al.* (2016) Simultaneous detection of mutations and copy number variation of NPM1 in the acute myeloid leukemia using multiplex ligation-dependent probe amplification. *Mutat. Res.*, **786**, 14–26.
39. 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. *et al.* (2013) Double nicking by RNA-Guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, **154**, 1380–1389.
40. Fu,Y.F., Foden,J.A., Khayter,C., Maeder,M.L., Reyon,D., Joung,J.K. and Sander,J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.*, **31**, 822–826.
41. Yang,Z., Steentoft,C., Hauge,C., Hansen,L., Thomsen,A.L., Niola,F., Vester-Christensen,M.B., Frodin,M., Clausen,H., Wandall,H.H. *et al.* (2015) Fast and sensitive detection of indels induced by precise gene targeting. *Nucleic Acids Res.*, **43**, e59.
42. Findlay,S.D., Vincent,K.M., Berman,J.R. and Postovit,L.M. (2016) A Digital PCR-Based method for efficient and highly specific screening of genome edited cells. *PLoS One*, **11**, e0153901.
43. Gaudelli,N.M., Komor,A.C., Rees,H.A., Packer,M.S., Badran,A.H., Bryson,D.I. and Liu,D.R. (2017) Programmable base editing of A.T to G.C in genomic DNA without DNA cleavage. *Nature*, **551**, 464–471.
44. Sentmanat,M.F., Peters,S.T., Florian,C.P., Connelly,J.P. and Pruett-Miller,S.M. (2018) A survey of validation strategies for CRISPR-Cas9 editing. *Sci. Rep.*, **8**, 888.
45. Mock,U., Hauber,I. and Fehse,B. (2016) Digital PCR to assess gene-editing frequencies (GEF-dPCR) mediated by designer nucleases. *Nat. Protoc.*, **11**, 598–615.
46. Hellmann,I., Lim,S.Y., Gelman,R.S. and Letvin,N.L. (2011) Association of activating KIR copy number variation of NK cells with containment of SIV replication in rhesus monkeys. *PLoS Pathog.*, **7**, e1002436.
47. Armour,J.A.L., Palla,R., Zeeuwen,P.L.J.M., den Heijer,M., Schalkwijk,J. and Hollox,E.J. (2007) Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. *Nucleic Acids Res.*, **35**, e19.
48. Marcinkowska-Swojak,M., Klonowska,K., Figlerowicz,M. and Kozlowski,P. (2014) An MLPA-based approach for high-resolution genotyping of disease-related multi-allelic CNVs. *Gene*, **546**, 257–262.
49. Imbard,A., Pasmant,E., Sabbagh,A., Luscan,A., Soares,M., Goussard,P., Blanche,H., Laurendeau,I., Ferkal,S., Vidaud,M. *et al.* (2015) NF1 single and multi-exons copy number variations in neurofibromatosis type 1. *J. Hum. Genet.*, **60**, 221–224.
50. Gatta,V., Scarciolla,O., Gaspari,A.R., Palka,C., De Angelis,M.V., Di Muzio,A., Guanciali-Franchi,P., Calabrese,G., Uncini,A. and Stuppia,L. (2005) Identification of deletions and duplications of the DMD gene in affected males and carrier females by multiple ligation probe amplification (MLPA). *Hum. Genet.*, **117**, 92–98.



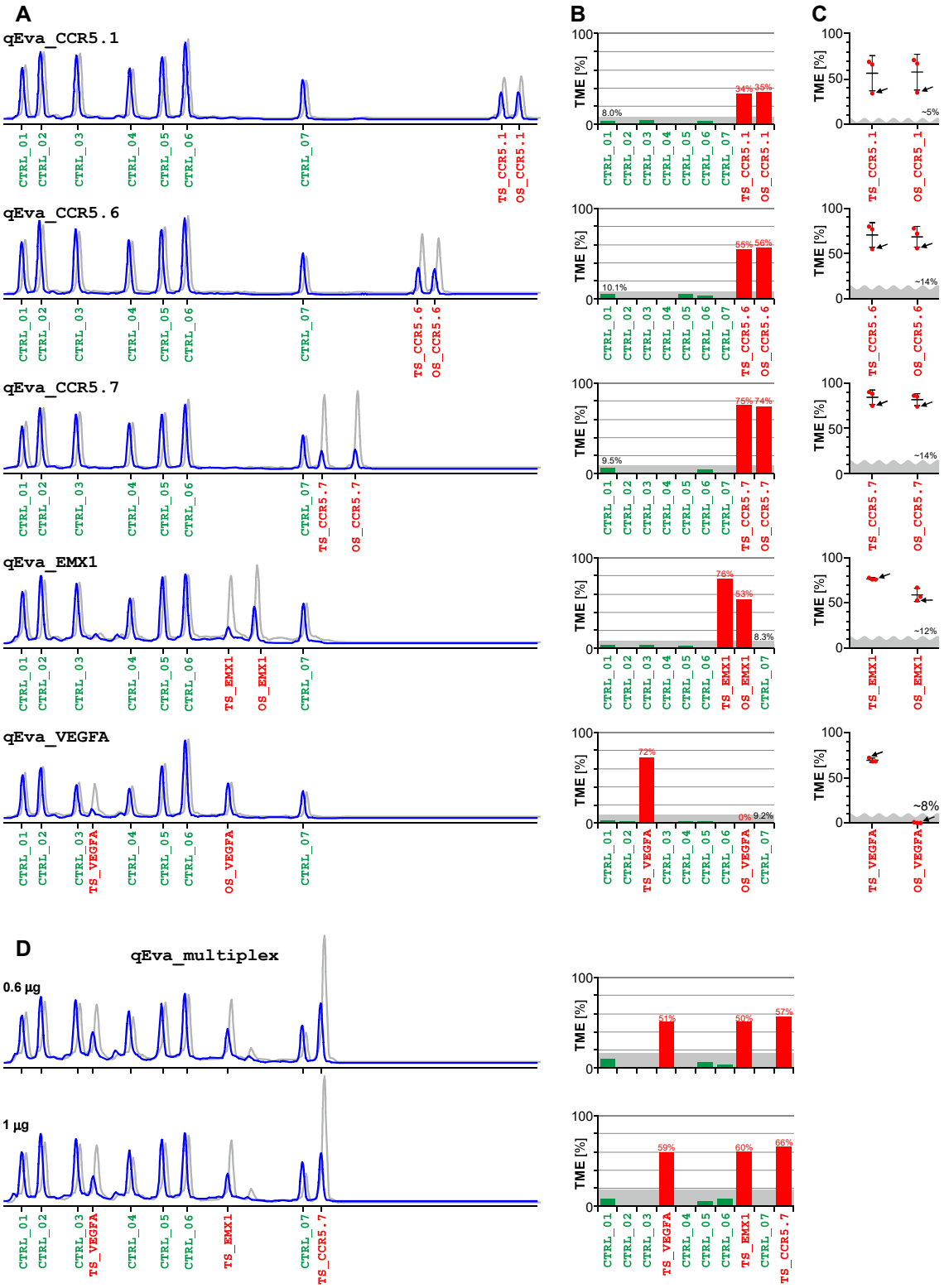
Supplementary Figure S1. Analysis of the target (*TP53*) and off-target mutation efficiencies in HCT116 cells that were either transfected (1.5 µg of plasmid DNA) or not transfected with the Cas9_E4.1 plasmid. The figure scheme is the same as the one used in Figure 1B.



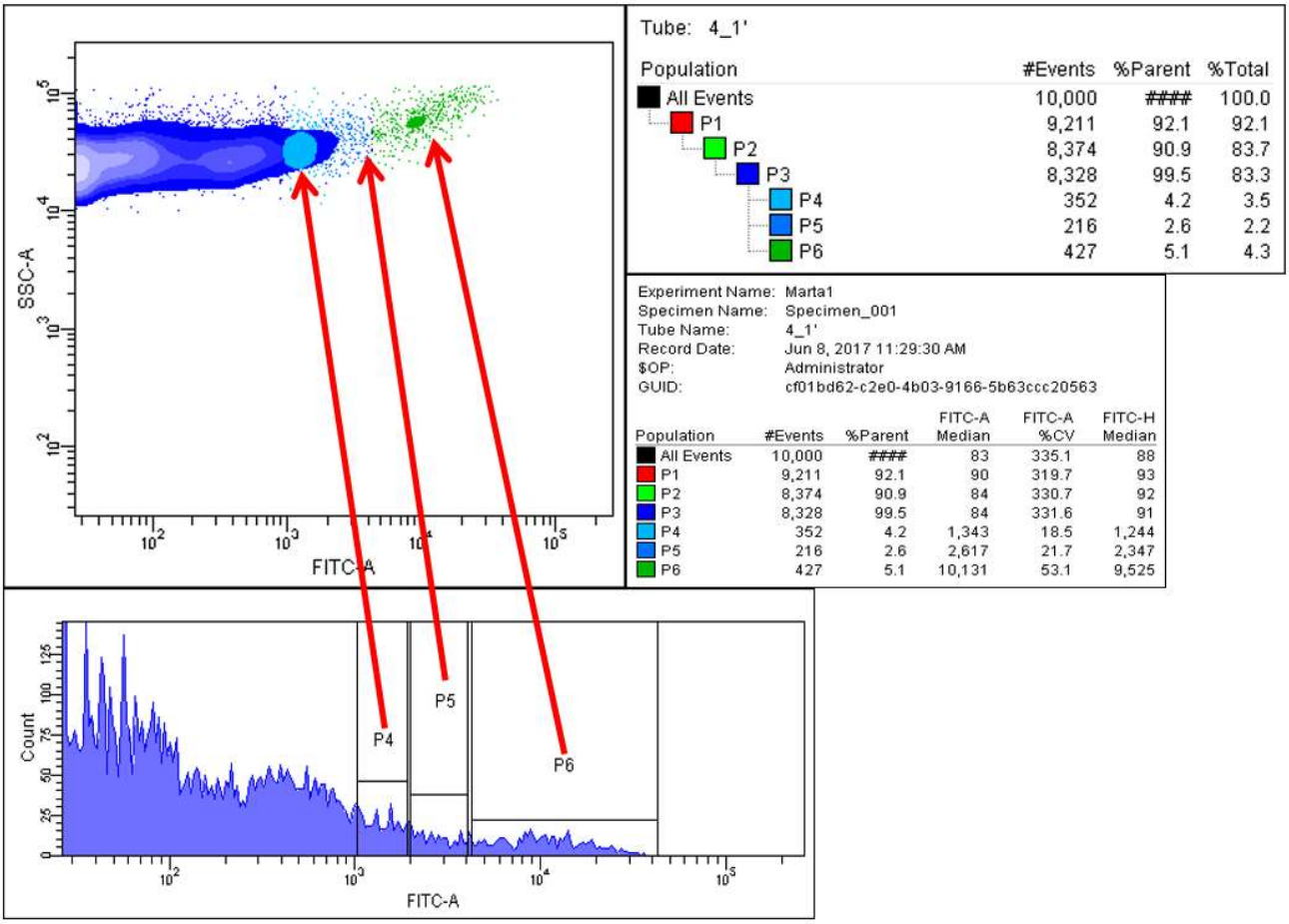
Supplementary Figure S2. The sensitivity of MLPA probes utilized in the qEva-CRISPR assays to detect single-nucleotide mutations at the predicted target-cleavage-sites. Analysis of a reference (Cas9-untreated) DNA sample using the qEva_E4.1 assay with a target-specific probe: TS_E4.1, which perfectly matched the target sequence (upper panel); TS_E4.1_A/T, which included a single-nucleotide mismatch directly adjacent to the ligation position (middle panel); and TS_E4.1_delA, which included a single-nucleotide deletion directly adjacent to the ligation position (lower panel). The nucleotide modifications were introduced in the last 3' nucleotide of the 5' half-probe. The 3'half-probe remained unchanged in all of the target-specific probes. From left to right: (i) overlapped electropherograms of the analyzed (blue) and reference (gray) sample, (ii) RS plot, and (iii) 5' and 3' RSSs of the target-specific probes used in the experiment (red font) along the target sequence (gray font); the position of the introduced sequence modification is indicated (yellow background). The general electropherogram and chart designations are the same as those described in Figure 1B. Red arrowheads indicate the expected position of target-specific probe signals. As shown, single-nucleotide mutations prevent half-probe ligations and completely abrogate the probe signal.



Supplementary Figure S3. Specificity and sensitivity of qEva-CRISPR/MLPA probes tested with the use of common SNPs. **(A)** Multiplex genotyping of three SNPs (i.e., SNP1_C/G rs11766125, SNP2_delC rs150749128, and SNP3_insT rs11396214) in ten reference HapMap samples from European (CEU), Asiatic-Japan (JPT), and African-Yoruba (YPT) populations (blue electropherograms). The SNP genotyping assay included three probes specific to the reference sequences (alleles) of the selected SNPs mixed with 8 control probes (for details, see Supplementary Table S3). As shown, signals (peaks height) of the SNP-specific probes correlate perfectly with SNP genotypes determined in the HapMap samples (as indicated over the corresponding peaks; +/+ homozygote for a reference allele, +/- heterozygote, and -/- homozygote for an alternative allele). The gray-electropherogram overlapped with each blue-electropherogram represents the hypothetical sample homozygous (+/+) for each of the tested SNPs. **(B)** Relative signals of the tested SNPs in the homozygous (+/+ and -/-), heterozygous (+/-) and serially diluted DNA samples: (left-hand side) the corresponding fragments of the electropherograms show the signal of the tested SNPs (the calculated relative signal value is indicated in red) along with the signal of the closest control probe (the calculated relative signal values are indicated in red), and (right-hand side) scatter plots showing the correlation between the calculated relative signals (x-axes) and the dilution factors of the samples (y-axes). Correlation coefficients (R2) are indicated on the graphs, ND – signal not detected. The electropherograms scheme is the same as the scheme in Figure 1B.



Supplementary Figure S4. qEva-CRISPR analysis of target modifications induced by well-validated sgRNAs. **(A)-(C)** qEva-CRISPR analysis of 5 different targets and corresponding off-targets located in 3 different genes, i.e., CCR5, EMX1, and VEGFA. Representative electropherograms **(A)** and corresponding TME bar plots **(B)** of samples treated with particular CRISPR/Cas9 reagents. **(C)** Dot plots summarizing results of three independent experiments. Dots represent TME values of targets and corresponding off-targets. Horizontal lines represent medians, and whiskers represent SD values. Arrowheads indicate the TME values corresponding to results shown in A and B. Note that the targets located in CCR5 are very similar [identical (TS_CCR5.1) or differ by a single nt substitution located >5 nt from PAM (TS_CCR5.6 and TS_CCR5.7)] and were similarly modified to the corresponding off-targets (Supplementary Table S2). To distinguish the identical sequences of TS_CCR5.1 and OT_CCR5.1, we took advantage of a previously described strategy using three-oligonucleotide probes (Serizawa RR, J Mol Diagn 2010; Langerak P, NAR 2005). **(D)** Multiplex analysis (electropherograms and TME-bar-plots) of three targets located in three different genes that were simultaneously modified by CRISPR/Cas9 reagents (plasmids) transfected under two different conditions (0.6 µg and 1 µg of the total amount of three plasmids). The electropherograms and TME bar plots schemes are the same as the ones used in Figure 1B.



Supplementary Figure S5. The cell sorting (gating) used to select three fractions of GFP-positive cells. The sorting was performed with the use of a BD FACS Aria™III (BD Biosciences) flow cytometer and analyzed with FACS DIVA software. For details, see Materials and Methods. P4, P5, and P6 correspond to low, medium, and high fluorescence intensity fractions, respectively.

Supplementary Table S1. Oligonucleotides used in the study

Oligonucleotide ID	Sequence (5'-3')	Description
P53E4-1s	CACCGGTGCAGGGGCCGCCGGTGT	oligo for Cas9_E4.1 plasmid construction
P53E4-1a	AAACACACCGGCGGCCCCTGCACC	oligo for Cas9_E4.1 plasmid construction
P53E4-2s	CACCGGGCAGCTACGGTTTCCGTCT	oligo for Cas9_E4.2 plasmid construction
P53E4-2a	AAACAGACGGAAACCGTAGCTGCCC	oligo for Cas9_E4.2 plasmid construction
VEGFA2s	CACCGACCCCCTCCACCCCGCCTC	oligo for Cas9_VEGFA plasmid construction
VEGFA2a	AAACGAGGCGGGGTGGAGGGGGTC	oligo for Cas9_VEGFA plasmid construction
EMX1s	CACCGAGTCCGAGCAGAAGAAGAA	oligo for Cas9_EMX1 plasmid construction
EMX1a	AAACTTCTTCTTCTGCTCGGACTC	oligo for Cas9_EMX1 plasmid construction
CCR5R-25s	CACCGTGTTCATCTTTGGTTTTGT	oligo for Cas9_CCR5.1 plasmid construction
CCR5R-25a	AAACACAAAACCAAAGATGAACAC	oligo for Cas9_CCR5.1 plasmid construction
CCR5#6s	CACCATGAACACCAGTGAGTAGAG	oligo for Cas9_CCR5.6 plasmid construction
CCR5#6a	AAACCTCTACTCACTGGTGTTCATC	oligo for Cas9_CCR5.6 plasmid construction
*CCR*5#7s	CACCGAACACCAGTGAGTAGAGCGG	oligo for Cas9_CCR5.7 plasmid construction
*CCR*5#7a	AAACCCGCTCTACTCACTGGTGTTC	oligo for Cas9_CCR5.7 plasmid construction
sgRNA1s	CACCGCTGCTGCTGCTGCTGCTGGA	oligo for Cas9(n)_HTT.sg1 plasmid construction
sgRNA1a	AAACTCCAGCAGCAGCAGCAGCAGC	oligo for Cas9(n)_HTT.sg1 plasmid construction
sgRNA3s	CACCGGAAGGACTTGAGGGACTCGA	oligo for Cas9_HTT.sg3 plasmid construction
sgRNA3a	AAACTCGAGTCCCTCAAGTCCTTCC	oligo for Cas9_HTT.sg3 plasmid construction
sgRNA4s	CACCGGCTTCCTCAGCCGCCGCCGC	oligo for Cas9(n)_HTT.sg4 plasmid construction
sgRNA4a	AAACGCGGCGGCGGCTGAGGAAGCC	oligo for Cas9(n)_HTT.sg4 plasmid construction
R_1s	ATAGGCTGCTGCTGCTGCTGCTGGA	oligo for p31 plasmid construction (T7 *in vitro* transcription)

R_1a	AAACTCCAGCAGCAGCAGCAGC	oligo for p31 plasmid construction (T7 *in vitro* transcription)
MK024	TGGTCCTCTGACTGCTCTT	forward primer for T7E1 analysis of *TP53* gene
MK025	GGTGAAGAGGAATCCCAAAGT	reverse primer for T7E1 analysis of *TP53* gene
U6-Fwd	GAGGGCCTATTTCCCATGATTCC	sequencing primer for sgRNA validation
WSF6	CAGCGTTTCTGGGTGAGC	sequencing primer for transcript validation
T7	TAATACGACTCACTATAGGG	sequencing primer for TP53_E4.1 and E4.2 single-cell-derived clones validation
ssODN	GACCGCCATGGCGACCCTGGAAAAGCTGATGA AGGCCTTCGAGTCCCTCAAGTCCTTCCAATGG TGAGCAAGGGCGAAAGCTTGTTCACCGGGGTG GTGCCCATCCTGGTCGAGCTGGACGCAGGCAC AGCCGCTGCTGCCTCAGCCGCAGCCGCCCCCG CCGCCGCCCCCGCCGCCAC	donor template for HDR repair

Supplementary Table S2. Target and off-target sequences analyzed with the use of the qEva-CRISPR assays

qEva-CRISPR assay ID	target/off-target probe ID	target/off-target ID	**target**/off-target, PAM sequence	mismatch #	target/off-target position (hg38)
qEva_E4.1	**TS_E4.1**	**E4.1**	**GGTGCAGGGGCCGCCGGTGT AGG**	**-**	**chr17:7676109-7676128(+)**
	OS_E4.1_1	E4.1_1	GGTGCTGGGGCTGCTGGTGT GGG	3	chr16:57262052-57262071(-)
	OS_E4.1_2	E4.1_2	GGTGCTGGGGCTGCCAGTGG TGG	4	chrX:45381921-45381940(-)
qEva_E4.2	**TS_E4.2**	**E4.2**	**GGCAGCTACGGTTTCCGTCT GGG**	**-**	**chr17:7676037-7676056(-)**
	OS_E4.2_1	E4.2_1	GCCAGCTAATGCTTCCGTCT TGG	4	chr3:40477139-40477158(+)
	OS_E4.2_2	E4.2_2	GGCAGATATGTTTTCTGTCT TGG	4	chr12:68858988-68859007(+)
	OS_E4.2_3	E4.2_3	GGCAGGTACCGTTTCCTTCC TGG	4	chr1:234478164-234478183(+)
qEva_HTT.sg1	**TS_HTT.sg1**	**HTT.sg1**	**CTGCTGCTGCTGCTGCTGGA AGG**	**-**	**chr4:3074875-3074894(-)**
	OS_HTT.sg1_1	HTT.sg1_1	CTGCTGCTGCTGCTGCTGCA GGG	1	chr11:117295639-117295658(+)
	OS_HTT.sg1_2	HTT.sg1_2	CTGCTGCTGCTGCTGCTGCA AGG	1	chr15:47717511-47717530(+)
qEva_HTT.sg3	**TS_HTT.sg3**	**HTT.sg3**	**GAAGGACTTGAGGGACTCGA AGG**	**-**	**chr4:3074857-3074876(-)**
	OS_HTT.sg3_1	HTT.sg3_1	GAAGGACTGCAGGTACTCGA AGG	3	chr1:27612971-27612990(+)
	OS_HTT.sg3_2	HTT.sg3_2	GATGGACTTGAGGGCCTCTA AGG	3	chr9:128637123-128637142(-)
qEva_HTT.sg4	**TS_HTT.sg4**	**HTT.sg4**	**GCTTCCTCAGCCGCCGCCGC AGG**	**-**	**chr4:3074975-3074994(+)**
	OS_HTT.sg4_1	HTT.sg4_1	GCTTCCGGAGCCGCCGCCGC AGG	2	chr16:66402674-66402693(-)
qEva_VEGFA	**TS_VEGFA**	**VEGFA**	**GACCCCCTCCACCCCGCCTC AGG**	**-**	**chr6:43,770,822-43,770,841(-)**
	OS_VEGFA	VEGFA_1	CTCCCCACCCACCCCGCCTC AGG	4	chr4:38,535,990-38,536,009(+)
qEva_EMX1	**TS_EMX1**	**EMX1**	**GAGTCCGAGCAGAAGAAGAA AGG**	**-**	**chr2:72,933,853-72,933,872(+)**
	OS_EMX1	EMX1_1	GAGTTAGAGCAGAAGAAGAA AGG	2	chr5:45,358,962-45,358,981(-)
qEva_CCR5.1	**TS_CCR5.1**	**CCR5.1**	**GTGTTCATCTTTGGTTTTGT GGG**	**-**	**chr3:46,373,020-46,373,039(+)**
	OS_CCR5.1	CCR5.1_1	GTGTTCATCTTTGGTTTTGT GGG	0	chr3:46,357,681-46,357,700(+)
qEva_CCR5.6	**TS_CCR5.6**	**CCR5.6**	**ATGAACACCAGTGAGTAGAG CGG**	**-**	**chr3:46,373,008-46,373,027(-)**
	OS_CCR5.6	CCR5.6_1	ATGAACACCAGCGAGTAGAG CGG	1	chr3:46,357,669-46,357,688(-)
qEva_CCR5.7	**TS_CCR5.7**	**CCR5.7**	**AACACCAGTGAGTAGAGCGG AGG**	**-**	**chr3:46,373,005-46,373,024(-)**
	OS_CCR5.7	CCR5.7_1	AACACCAGCGAGTAGAGCGG AGG	1	chr3:46,357,666-46,357,685(-)

red font – mismatches between the target and off-target sequences

Supplementary Table S3. Detailed characteristics and sequences of the probes included in the qEva-CRISPR assays

				5'half-probe (5'HP)								linker probe (LP)			3'half-probe (3'HP)							
probe id	gene	probe location (hg38)	comments	5'PSS	length	5'SS	length	5'TSS	length	Tm	5'HPL	LS	length	Tm	3'TSS	length	Tm	3'SS	length	3'PSS	length	3'HPL	TPL
qEva_E4.1																							
TS_E4.1	TP53	chr17:7,676,105-7,676,146		GGGTTCCCTAAGGGTTGGA	19	cgctactacta	11	CCTGCACAGCAGCTCCTACA	21	71,4	51	-	-	-	CCGGCGGCCCTGCACCAGCC	21	86,9	aatctac	7	TCTAGATTGGATCTTGCTGGCGC	23	51	102
OS_E4.1_1	PLLP	chr16:57,262,034-57,262,075		GGGTTCCCTAAGGGTTGGA	19	cgctactactattagta	17	TGTCCGCAATCAGCGCCACA	21	75,5	57	-	-	-	CCAGCAGCCCCAGCACCTAGG	21	76,5	aaactaaatctac	13	TCTAGATTGGATCTTGCTGGCGC	23	57	114
OS_E4.1_2	KDM5C	chrX:53,193,089-53,193,130		GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaa	20	CCCCTGCACCCATCCCCACCA	21	79,9	60	-	-	-	CTGGCAGCCCCAGCACCCAGG	21	80,9	gtcaaactaaatctac	14	TCTAGATTGGATCTTGCTGGCGC	23	60	120
qEva_E4.1_delA																							
TS_E4.1_delA	TP53	chr17:7,676,105-7,676,146		GGGTTCCCTAAGGGTTGGA	19	cgctactacta	11	CCTGCACAGCAGCTCCTAC	21	70,1	50	-	-	-	CCGGCGGCCCTGCACCAGCC	21	86,9	aatctac	7	TCTAGATTGGATCTTGCTGGCGC	23	51	101
qEva_E4.1_A/T																							
TS_E4.1_A/T	TP53	chr17:7,676,105-7,676,146		GGGTTCCCTAAGGGTTGGA	19	cgctactacta	11	CCTGCACAGCAGCTCCTACT	21	70,4	51	-	-	-	CCGGCGGCCCTGCACCAGCC	21	86,9	aatctac	7	TCTAGATTGGATCTTGCTGGCGC	23	51	102
qEva_E4.2																							
TS_E4.2	TP53	chr17:7,676,017-7,676,060		GGGTTCCCTAAGGGTTGGA	19	cgctactacta	11	CCAGGGCAGCTACGGTTTCCG	21	73	51	-	-	-	TCTGGGCTTCTTGCATTCTGGGA	23	71,4	tctac	5	TCTAGATTGGATCTTGCTGGCGC	23	51	102
OS_E4.2_1	ZNF619	chr3:40,477,135-40,477,176		GGGTTCCCTAAGGGTTGGA	19	cgctactactattagta	17	GCGCTCGCAAGGTGGCCAAGA	21	77,2	57	-	-	-	CGGAAGCATTAGCTGGCCGGC	21	75,7	aaactaaatctac	13	TCTAGATTGGATCTTGCTGGCGC	23	57	114
OS_E4.2_2	CPM	chr12:68,858,974-68,859,035		GGGTTCCCTAAGGGTTGGA	19	cgctacta	8	CATTACCCAATGTAATTGTGGAAG TCCAAGA	31	71,6	58	-	-	-	CAGAAAACATATCTGCCCTATAG AACCAAC	31	71	tctac	5	TCTAGATTGGATCTTGCTGGCGC	23	59	117
OS_E4.2_3	TARBP1	chr1:234,478,157-234,478,201		GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaa	20	CTGCACCTGCGGGCCCCAGGA	21	82,1	60	-	-	-	AGGAAACGGTACCTGCCTCTTTGC	24	72,3	aaactaaatctac	13	TCTAGATTGGATCTTGCTGGCGC	23	60	120
qEva_HTT.sg1																							
TS_HTT.sg1	HTT	chr4:3,074,855-3,074,899		GGGTTCCCTAAGGGTTGGA	19	cgctactact	10	TGCTGCTGCTGCTGCTGCTGCT	22	75,1	51	-	-	-	GGAAGGACTTGAGGGACTCGAAG	23	71	tctac	5	TCTAGATTGGATCTTGCTGGCGC	23	51	102
OS_HTT.sg1_1	BACE1	chr11:117,295,634-117,295,675		GGGTTCCCTAAGGGTTGGA	19	cgctactactattagt	16	TGTGCCTGCTGCTGCTGCTGCT	22	75,2	57	-	-	-	GCAGGGACAGCTAGTGGGCA	20	71,8	caaactaaatctac	14	TCTAGATTGGATCTTGCTGGCGC	23	57	114
OS_HTT.sg1_2	SEMA6D	chr15:47,717,508-47,717,547		GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtaga	19	AACTCGCCACCACCCCTTGC	20	72,7	58	-	-	-	AGCAGCAGCAGCAGCAGCGG	20	74,5	gtcaaactaaatctac	14	TCTAGATTGGATCTTGCTGGCGC	23	59	117
qEva_HTT.sg3																							
TS_HTT.sg3	HTT	chr4:3,074,837-3,074,883		GGGTTCCCTAAGGGTTGGA	19	cgctacta	8	GCTGCTGGAAGGACTTGAGGACT	23	73,8	51	-	-	-	CGAAGGCCTTCATCAGCTTTCC	23	71,2	tctac	5	TCTAGATTGGATCTTGCTGGCGC	23	51	102
OS_HTT.sg3_1	RGR	chr1:27,612,964-27,613,008		GGGTTCCCTAAGGGTTGGA	19	cgctactactatta	14	CCTCCAGGAAGGACTGCAGGTACT	24	72,5	57	-	-	-	CGAAGGTAGGCCTCTCCTCCG	21	71	aaactaaatctac	13	TCTAGATTGGATCTTGCTGGCGC	23	57	114
OS_HTT.sg3_2	MDR34	chr9:128,637,105-128,637,146		GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtag	18	GGCTGATGGACTTGAGGGCCT	21	72,6	58	-	-	-	CTAGGGCTAAGGGTGGCCTGG	21	72,9	tcaaactaaatctac	14	TCTAGATTGGATCTTGCTGGCGC	23	59	117
qEva_HTT.sg4																							
TS_HTT.sg4	HTT	chr4:3,074,972-3,075,011		GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	TCAGCTTCCTCAGCCGCCGC	20	75,2	51	-	-	-	CGCAGGCACAGCCGCTGCTG	20	78,8	aaatctac	8	TCTAGATTGGATCTTGCTGGCGC	23	51	102
OS_HTT.sg4_1	CDH5	chr16:66,402,656-66,402,696		GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtag	18	CCTGCTTCCGGAGCCGCCGC	20	81,3	57	-	-	-	CGCAGGAAGATGAGCAGGGTG	21	72,5	aaactaaatctac	13	TCTAGATTGGATCTTGCTGGCGC	23	57	114
qEva_HTT.edit																							
TS_HTT.sg4	HTT	chr4:3,074,972-3,075,011	as in qEva_HTT.sg4	GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	TCAGCTTCCTCAGCCGCCGC	20	75,2	51	-	-	-	CGCAGGCACAGCCGCTGCTG	20	78,8	aaatctac	8	TCTAGATTGGATCTTGCTGGCGC	23	51	102
TS_HTT.sg1/4	HTT	chr4:3,074,855-3,074,898	composed of 5'HP of TS_HTT.sg1* and 3'HP of TS_HTT.sg4	GGGTTCCCTAAGGGTTGGA	19	cgctactacta	11	CTTCGAGTCCCTCAAGTCCTTCCA	24	72,3	54	-	-	-	CGCAGGCACAGCCGCTGCTG	20	78,8	aaatctac	8	TCTAGATTGGATCTTGCTGGCGC	23	51	105
TS_HTT.sg1*	HTT	chr4:3,074,855-3,074,898	specific to the other strand than TS_HTT.sg1	GGGTTCCCTAAGGGTTGGA	19	cgctactacta	11	CTTCGAGTCCCTCAAGTCCTTCCA	24	72,3	54	-	-	-	GCAGCAGCAGCAGCAGCAGC	20	74,9	actaaatctac	11	TCTAGATTGGATCTTGCTGGCGC	23	54	108
TS_HTT.HDR	HTT	chr4:3,074,799-3,074,878	three-oligonucleotide probe	GGGTTCCCTAAGGGTTGGA	19	-	0	GCCCTTGCTCACCATTGGAAG	21	71	40	GACTTGAGGGACTCGAAGGCCTTC ATCAGCTTTTCCAGGGTGGCCATG GCGGTC	54	92,3	TCCCGCCCGGCACGGCAGTC	20	81,4	-	0	TCTAGATTGGATCTTGCTGGCGC	23	43	137
qEva_SNP																							
SNP1_C/G	MUC12	chr7:101,004,811-101,004,859		GGGTTCCCTAAGGGTTGGA	19	cgctactactatt	13	CAAACACTCTCACCTGCCAGCATG A	25	73,5	57	-	-	-	CAAGCTCCAGCATCAGTGGAGAAC	24	71,7	ctaaatctac	10	TCTAGATTGGATCTTGCTGGCGC	23	57	114
SNP2_delC	ADCY2	chr5:7,681,102-7,681,152		GGGTTCCCTAAGGGTTGGA	19	cgctactactatta	14	CTTTCCGCTTTCTGATGTGATGAC TTC	27	70,8	60	-	-	-	CCTGCAGGTTGCCTTTTGCATCTTT	24	70,4	aaactaaatctac	13	TCTAGATTGGATCTTGCTGGCGC	23	60	120
SNP3_insT	EGFR	chr7:55,066,427-55,066,469		GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaatt	22	GGGTTGTAGCCTCGAATTCTCG	23	71,6	64	-	-	-	GGCCAGGGCTACCGTCAGCC	20	76,7	taatggtcaaactaaatctac	21	TCTAGATTGGATCTTGCTGGCGC	23	64	128
qEva_VEGFA																							
TS_VEGFA	VEGFA	chr6:43,770,805-43,770,844		GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	CCCGACCCCTCCACCCCGC	20	84,3	51	-	-	-	CTCCGGGCGCGGGCTCCGGC	20	86,5	aaatctac	8	TCTAGATTGGATCTTGCTGGCGC	23	51	102
OS_VEGFA	intergenic	chr4:38,535,986-38,536,032		GGGTTCCCTAAGGGTTGGA	19	cgctactactattagta	17	TCCCCTCCCCACCCACCCGC	21	83,6	57	-	-	-	CTCAGGCTTGAAGAGGAAAGAAGA GC	26	71,1	aaatctac	8	TCTAGATTGGATCTTGCTGGCGC	23	57	114
qEva_EMX1																							
TS_EMX1	EMX1	chr2:72,933,848-72,933,892		GGGTTCCCTAAGGGTTGGA	19	cgctactactattagta	17	GGCCTGAGTCCGAGCAGAAGAA	22	72,4	58	-	-	-	GAAGGGCTCCCATCACATCAACC	23	72,9	aaactaaatctac	13	TCTAGATTGGATCTTGCTGGCGC	23	59	117
OS_EMX1	NCN	chr5:45,358,937-45,358,996		GGGTTCCCTAAGGGTTGGA	19	cgctactac	9	CAAGACAGATTGTCAGAGTTAGAG CAGAAGAA	32	71,6	60	-	-	-	GAAAGGCATGGAGTAAAGGCAATC TTGT	28	71,2	taaatctac	9	TCTAGATTGGATCTTGCTGGCGC	23	60	120
qEva_CCR5.1																							
TS_CCR5.1	CCR5	chr3:46,373,007-46,373,085	three-oligonucleotide probe; 5'HP common with OS_CCR5.1	GGGTTCCCTAAGGGTTGGA	19	cgctactactattag	15	GCTCTACTCACTGGTGTTCATCTT TGGTTT	30	71,6	64	TGTGGGCAACATGCTGGTC	19	67,8	ATCCTCATCCTGATAAACTGCAAA AGGCTG	30	72,8	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	65	148
OS_CCR5.1	CCR2	chr3:46,357,669-46,357,751	three-oligonucleotide probe; 5'HP common with TS_CCR5.1	GGGTTCCCTAAGGGTTGGA	19	cgctactactattag	15	GCTCTACTCACTGGTGTTCATCTT TGGTTT	30	71,6	64	TGTGGGCAACATGCTGGTC	19	67,8	GTCCTCATCTTAATAAACTGCAAA AAGCTGAAGTG	35	72,0	taaatctac	9	TCTAGATTGGATCTTGCTGGCGC	23	67	150
qEva_CCR5.6																							
TS_CCR5.6	CCR5	chr3:46,372,995-46,373,042	3'HP common with OS_CCR5.6	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattga	24	CCCACAAAACCAAAGATGAACACC AGT	27	72,3	70	-	-	-	GAGTAGAGCGGAGGCAGGAGG	21	71,9	gtatctaatggtcaaactaaatct ac	24	TCTAGATTGGATCTTGCTGGCGC	23	70	140
OS_CCR5.6	CCR2	chr3:46,357,654-46,357,702	3'HP common with TS_CCR5.6	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattga tgc	27	CCACAAAACCAAAGATGAACACCA GC	26	72,6	72	-	-	-	GAGTAGAGCGGAGGCAGGAGG	21	71,9	gtatctaatggtcaaactaaatct ac	24	TCTAGATTGGATCTTGCTGGCGC	23	70	142
qEva_CCR5.7																							
TS_CCR5.7	CCR5	chr3:46,372,988-46,373,038		GGGTTCCCTAAGGGTTGGA	19	cgctactactatta	14	CAAAACCAAAGATGAACACCAGTG AGTAGAG	31	71,6	64	-	-	-	CGGAGGCAGGAGGCGGGCTG	20	80,9	taatggtcaaactaaatctac	21	TCTAGATTGGATCTTGCTGGCGC	23	64	128
OS_CCR5.7	CCR2	chr3:46,357,649-46,357,697		GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaa	20	AAACCAAAGATGAACACCAGCGAG TAGAG	29	71,1	68	-	-	-	CGGAGGCAGGAGTTGGGCCC	20	78,4	tatctaatggtcaaactaaatcta c	25	TCTAGATTGGATCTTGCTGGCGC	23	68	136

qEva_multiplex																						
TS_VEGFA	VEGFA	chr6:43,770,805-43,770,844	as in qEva_VEGFA	GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	CCCGACCCCCTCCACCCGCC	20	84,3	51	-	-	-	CTCCGGGCGCGGGCTCCGGC	20	86,9	aaatctac	8	TCTAGATTGGATCTTGCTGGCGC	23	51	102
TS_EMX1	EMX1	chr2:72,933,848-72,933,892	as in qEva_EMX1	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagta	17	GGCCTGAGTCCGAGCAGAAGAA	22	72,4	58	-	-	-	GAAGGGCTCCCATCACATCAACC	23	72,9	aaactaaatctac	13	TCTAGATTGGATCTTGCTGGCGC	23	59	117
TS_CCR5.7	CCR5	chr3:46,372,988-46,373,038	as in qEva_CCR5.7	GGGTTCCCTAAGGGTTGGA	19	cgctactactatta	14	CAAAACCAAAGATGAACACCAGTG AGTAGAG	31	71,6	64	-	-	-	CGGAGGCAGGAGGCGGGCTG	20	80,9	taatggtcaaactaaatctac	21	TCTAGATTGGATCTTGCTGGCGC	23	64	128
control probes																						
CTRL_01	NF2	chr22:29,673,307-29,673,349		GGGTTCCCTAAGGGTTGGA	19	cgctac	6	GGCCCAGATCACCGAGGAGGA	21	75,6	46	-	-	-	GGCAAAACTTCTGGCCCAGAAG	22	71	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	47	93
CTRL_02	DROSHA	chr5:31,410,760-31,410,801		GGGTTCCCTAAGGGTTGGA	19	cgctacta	8	GCCAAGGTCTTGGTGCGAAGC	21	73,2	48	-	-	-	GCCACAGGCCTCTTGGTCTTG	21	72,3	ctac	4	TCTAGATTGGATCTTGCTGGCGC	23	48	96
CTRL_03	C9orf9	chr9:132,886,392-132,886,438		GGGTTCCCTAAGGGTTGGA	19	cgctact	7	AAATTCCAGCGTGGGCCACCCAT	23	74,7	49	-	-	-	ACTGAGAGGGAAACAGCAGGGACT	24	70,5	tac	3	TCTAGATTGGATCTTGCTGGCGC	23	50	99
CTRL_04	LOC101928103	chr2:214,810,835-214,810,888		GGGTTCCCTAAGGGTTGGA	19	cgctac	6	CTGGGACCTGGATAGACACTTGGA TAT	27	70,3	52	-	-	-	CAGCTAGAAAGCTACGACAGGAAA CCA	27	71,1	tac	3	TCTAGATTGGATCTTGCTGGCGC	23	53	105
CTRL_05	Intergenic	chr17:59,840,751-59,840,797		GGGTTCCCTAAGGGTTGGA	19	cgctactacta	11	GGCTGCTGCATACTGCTAAATGGC	24	71,7	54	-	-	-	ACCTCTGGGATTGGCCTACCTGG	23	73	aaatctac	8	TCTAGATTGGATCTTGCTGGCGC	23	54	108
CTRL_06	LMNA	chr1:156,136,027-156,136,071		GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	CAGCTGGACGAGTACCAGGAGCTT	24	72,8	55	-	-	-	CTGGACATCAAGCTGGCCCTG	21	72,7	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	56	111
CTRL_07	BARD1	chr2:214,745,101-214,745,157		GGGTTCCCTAAGGGTTGGA	19	cgctactactattagt	16	ACTCATGTTGTTGTTCCTGGTGAT GCA	27	71,6	62	-	-	-	GTTCAAAGTACCTTGAAGTGTATG CTTGGG	30	71,5	taaatctac	9	TCTAGATTGGATCTTGCTGGCGC	23	62	124
CTRL_08	Intergenic	chr6:71,404,869-71,404,934		GGGTTCCCTAAGGGTTGGA	19	cgctactactatta	14	CTTAGTACTTGGCATTCATATACC ACTGCTGAC	33	71,2	66	-	-	-	TAGAGCTTGAGTCAATCCACCAAA CTCAGCTAT	33	72,3	ctaaatctac	11	TCTAGATTGGATCTTGCTGGCGC	23	66	132

Legend:

5'HP, 3'HP - 5' and 3'half-probe

LP - linker-probe

LS - linker sequence

5'PSS, 3'PSS - 5' and 3' primer-specific sequence, respectively

5'SS, 3'SS - 5' and 3' stuffer sequence, respectively

5'TSS, 3'TSS - 5' and 3' target-specific sequence, respectively

Tm - melting temperature

5'HPL, 3'HPL - 5' and 3' half-probe length

TPL - total probe length

SALSA PCR Forward primer (Labeled): *GGGTTCCCTAAGGGTTGGA

SALSA PCR Reverse primer (Unlabeled): GTGCCAGCAAGATCCAATCTAGA

AC# V00604 Phage M13 genome

position: 3-99

5'-cgctactactattagtagaattgatgccacctttcagctcgcgccccaaatgaaaatatagctaaacaggttattgaccatttgcgaaatgtatctaatggtcaaactaaatctac-3'

International Journal of
Molecular Sciences

MDPI

Article

Generation of New Isogenic Models of Huntington's Disease Using CRISPR-Cas9 Technology

Magdalena Dabrowska [1], Agata Ciolak [2], Emilia Kozlowska [2], Agnieszka Fiszer [2] and Marta Olejniczak [1,*]

[1] Department of Genome Engineering, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland; mdabrowska@ibch.poznan.pl
[2] Department of Medical Biotechnology, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland; aluzna@ibch.poznan.pl (A.C.); emiliak@ibch.poznan.pl (E.K.); agnieszka.fiszer@ibch.poznan.pl (A.F.)
* Correspondence: marta.olejniczak@ibch.poznan.pl; Tel.: +48-61-852-8503

Received: 8 January 2020; Accepted: 5 March 2020; Published: 8 March 2020

Abstract: Huntington's disease (HD) is a fatal neurodegenerative disorder caused by the expansion of CAG repeats in exon 1 of the huntingtin gene (*HTT*). Despite its monogenic nature, HD pathogenesis is still not fully understood, and no effective therapy is available to patients. The development of new techniques such as genome engineering has generated new opportunities in the field of disease modeling and enabled the generation of isogenic models with the same genetic background. These models are very valuable for studying the pathogenesis of a disease and for drug screening. Here, we report the generation of a series of homozygous HEK 293T cell lines with different numbers of CAG repeats at the *HTT* locus and demonstrate their usefulness for testing therapeutic reagents. In addition, using the CRISPR-Cas9 system, we corrected the mutation in HD human induced pluripotent stem cells and generated a knock-out of the *HTT* gene, thus providing a comprehensive set of isogenic cell lines for HD investigation.

Keywords: genome editing; iPSCs; aberrant splicing; CAG repeats; Huntington's disease; CRISPR

1. Introduction

Huntington's disease (HD) is an incurable and progressive neurodegenerative disease caused by the expansion of CAG repeats in exon 1 of the huntingtin gene (*HTT*) [1]. Depending on the number of CAG repeats, *HTT* alleles are classified as normal (<27), intermediate (28–35), reduced penetrance (36–39) and fully penetrant mutant alleles (>40) [2]. The most severe form of HD, known as juvenile HD, is associated with a >60 CAG expansion and onset before 20 years of age. Because HD is inherited in an autosomal dominant manner, patients harbor a single copy of mutant *HTT* encoding a huntingtin protein containing a polyglutamine (polyQ) domain [3]. The pathogenesis of HD is not fully understood; however, toxic gain-of-function resulting from mutant huntingtin is considered the most prominent cause [4–6]. During the disease course, loss of neurons in the striatum and cortex is observed, accompanied by reactive gliosis and astrocytosis, leading to progressive movement abnormalities and dementia. Different cellular models are being applied to study the pathogenesis of the disease and to test therapeutic approaches for HD. These models include HEK 293T cells with exogenous expression of mutant *HTT* or a fragment thereof, patient-derived fibroblasts, human and rodent induced pluripotent stem cells (iPSCs), neural stem cells (NSCs) and postmitotic neurons [7]. Healthy cells of different origins are generally used as controls in these experiments, increasing the risk of inappropriate interpretation of the observed phenotypes. The influence of the genetic background on the disease phenotype is increasingly being proven in HD, especially in light of the identification of

genetic modifiers that affect the age of onset of HD in genome-wide association studies (GWAS) [8]. Subtle differences in DNA may influence phenomena such as somatic instability or disease onset.

Despite the monogenic nature of HD and the fact that its genetic basis was discovered more than two decades ago, we still do not understand all aspects of HD pathogenesis, such as the somatic instability of CAG repeats or the role of normal huntingtin in the adult brain and other tissues. This knowledge would be very important for the justification of allele-selective vs nonselective therapeutic approaches for this still incurable disease [9,10]. Rapidly developing genome editing tools such as CRISPR-Cas9 provide an opportunity to generate new isogenic models without background-related variability and to fill these knowledge gaps [11–15]. The CRISPR-Cas9 system uses the Cas9 nuclease and a small guide RNA (sgRNA) for the site-specific cleavage of a target sequence containing a protospacer-adjacent motif (PAM) [16]. Double-strand breaks (DSBs) are mainly repaired by nonhomologous end joining (NHEJ) or homology-directed repair (HDR) when the donor template is delivered. In this way, genes are being knocked-out or the mutation is repaired or introduced into specific *loci*.

Because human iPSCs (hiPSCs) can be differentiated into virtually every cell type, they serve as a valuable model for disease modeling and drug screening [7,17–19]. A few isogenic pairs of mutant/control hiPSC lines for HD have been established using traditional homologous recombination [20] and CRISPR-Cas9 technology [21,22]. The corrected cell lines were differentiated into neural progenitor cells (NPCs) and active neurons, and the reversal of HD-associated phenotypes was observed. However, one of these isogenic pairs contains 18/180 CAG repeats, which is an extremely long mutant variant [21], and the second pair contains 19/97 mixed CAG/CAA repeats [22], which does not reflect the typical sequence or length of the mutant tract in *HTT*. A panel of isogenic HD human embryonic stem cells (hESCs) with a more relevant repeat number (30, 45, 65 and 81 CAGs) was generated recently with the use of transcription activator-like effector nucleases (TALENs) [23]. These cell lines were differentiated into NPCs, neurons, hepatocytes and muscle cells. Transcriptomic and proteomic analyses identified cell-type and CAG-repeat length-dependent phenotypes.

Drug screening experiments are usually performed using exogenous reporter systems expressing *HTT* gene fragments in easy-to-transfect cells such as HEK 293 cells. This method is convenient and quantitative; however, it is simplified and "artificial" and does not include the potential influence of genomic context, promoter strength, or the full length gene. Endogenous models of HD such as patient-derived fibroblasts are sensitive to plasmid transfection and are therefore not useful for vector-based drug screening [24]. HEK 293 cells and hiPSCs can be expanded indefinitely and are popular cellular models for many studies. Using genome editing technology, the endogenous *HTT* locus can be modified by lengthening the CAG repeat tract in frame to produce mutant huntingtin [25].

Here, we used various approaches involving CRISPR-Cas9 technology to successfully generate new isogenic models of HD. A series of HEK 293T cell lines with different numbers of CAG repeats at the endogenous *HTT* locus was generated, and the usefulness of these lines in the testing of therapeutic reagents was demonstrated. In addition, isogenic controls for juvenile HD hiPSCs (19/109 CAG repeats) were generated. These cell lines exhibit a normal-length CAG tract (19/19 CAG) in *HTT*. A hiPSC line with a knock-out of the *HTT* gene was also obtained. These cell lines were characterized in detail and can be used as valuable models to study the pathogenesis and therapy of HD.

2. Results

2.1. Generation of HEK 293T-Based Models of HD

To generate a model that is useful for studying the influence of repeat tract length on various aspects of the pathogenesis and therapy of HD, we generated HEK 293T cell lines with different numbers of CAG repeats at the *HTT* locus. To induce DSBs, we used previously validated sgRNAs [26] and different strategies: (i) Cas9 nickase (Cas9n) and an sgRNA pair (HTT_sg1 and HTT_sg4) flanking the repeat tract, (ii) wtCas9 or Cas9n and one sgRNA (HTT_sg4) targeting the sequence downstream of the CAG tract, (iii) wtCas9 and sgRNA (HTT_sg3) targeting the sequence located upstream of the

repeat tract and (iv) a ribonucleoprotein (RNP) complex composed of the Cas9 protein and HTT_sg3 RNA. These strategies are summarized in Table S1. Only the last strategy resulted in frequent HDR and the generation of modified cell lines, which presented 41 CAG, 53 CAG or 84 CAG repeats (Figure 1A,B). Interestingly, biallelic modifications were more frequent than mono-allelic modifications. In heterozygous clones, the sequence of the mutant alleles was corrected without any scars (by HDR), whereas normal alleles contained indel mutations at a site of Cas9 cutting (resulting from NHEJ). We characterized the clonal cell lines by PCR amplification of the *HTT* gene region encompassing the CAG repeats (Figure 1C) and by analysis of huntingtin transcript and protein levels. Generally, the *HTT* transcript level measured by RT-qPCR with a primer pair located downstream of the CAG tract was correlated with the length of the CAG tract (Figure 1D). Proteins detected by western blotting migrated in polyacrylamide gels according to polyQ domain length, and the signal intensity was inversely correlated with the length of the polyQ tract, which is typically observed (Figure 1E). The sequences of the selected clones were confirmed by Sanger sequencing (Figure S1).

The generated HD models with increasing numbers of CAG tracts may serve as useful HD models for drug screening, especially using RNAi-based CAG-targeting strategies. Therefore, we transfected the edited HEK 293T cells with two previously validated siRNAs targeting either the exon 1 sequence under a non-allele-selective strategy (siHTT) [27] or the CAG tract under an allele-selective strategy (siRNA_A2) [28]. We analyzed the efficiency of silencing 48 h post-transfection by western blotting (Figure 1F, Figure S2). The level of the HTT protein was reduced by ~50% and ~30% in all models by siHTT and siRNA_A2, respectively (Figure 1F).

Pathogenic process in HD involves the production of smaller, N-terminal fragments of HTT [29] and some of this toxic products result from CAG repeat length–dependent aberrant splicing of exon 1 *HTT* [30,31]. To detect the early intron 1 transcripts, we used nonquantitative RT-PCR assays with three primer pairs spanning the exon 1–exon 2 junction, the exon 1–intron and intron 1 sequences (Figure 1G). The early intron 1 transcripts were elevated in HEK 293T cells which presented 41 CAG, 53 CAG or 84 CAG repeats compared to unmodified HEK 293T cells.

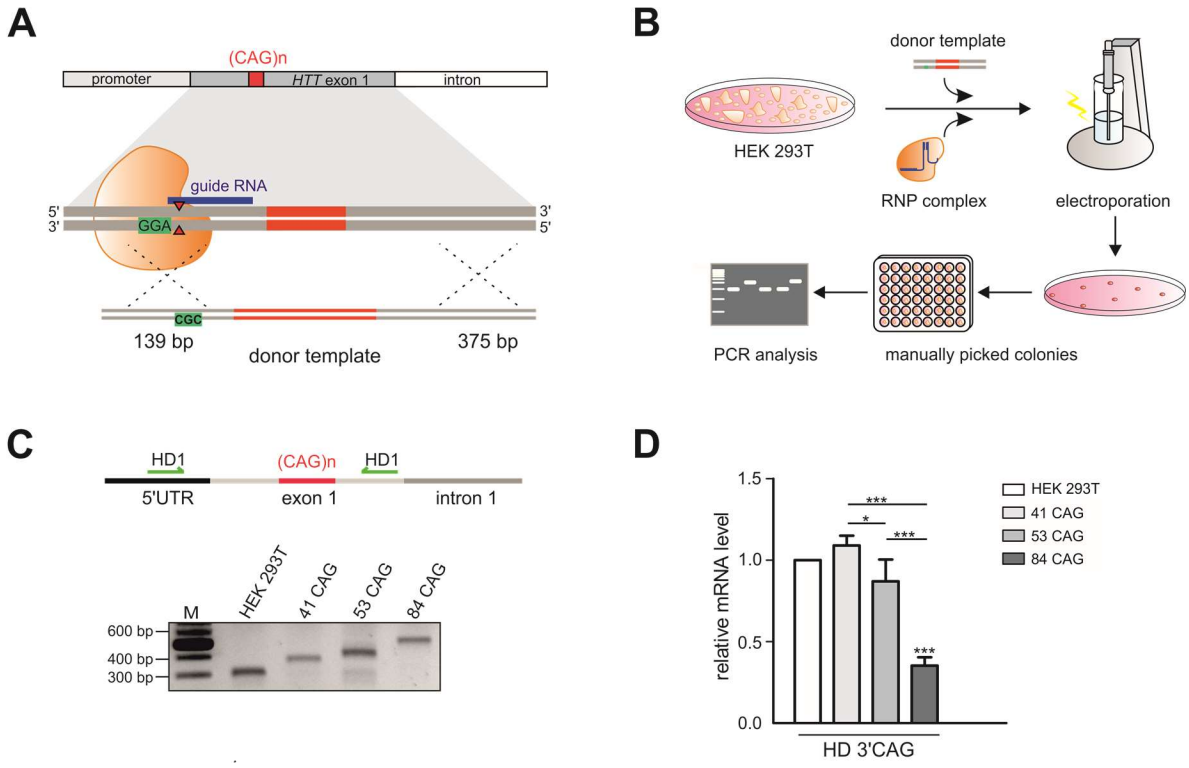


Figure 1. *Cont.*

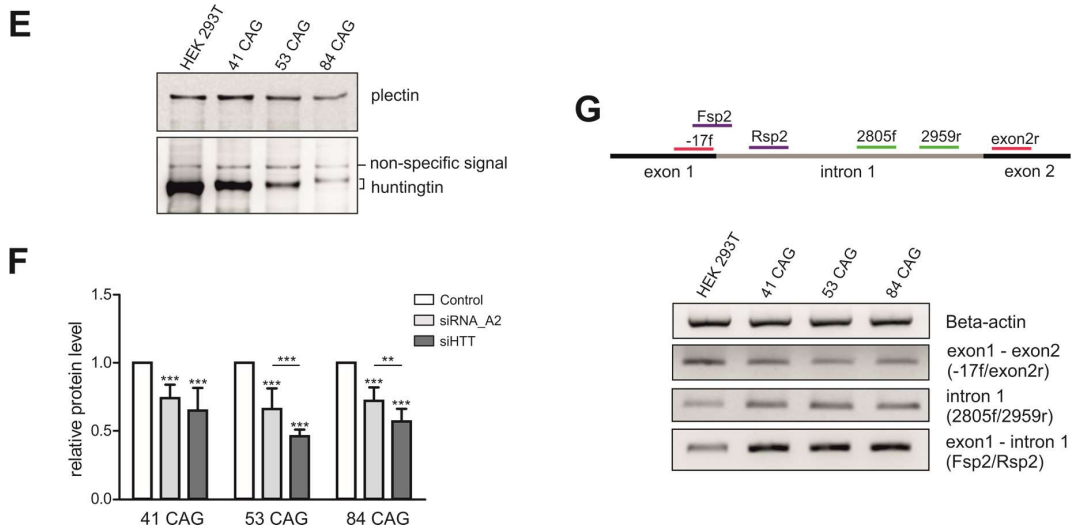


Figure 1. Generation of HEK 293T monoclonal cell lines with different numbers of CAG repeats at the *HTT locus*. (**A**) Successful strategy used to generate three monoclonal cell lines: 41 CAG, 53 CAG and 84 CAG. HTT_sg3 and SpCas9 were used to create DSB upstream from the CAG repeats. A linearized plasmid containing 41, 53 or 84 CAG repeats and silent mutation in a PAM sequence that doesn't lead to an amino acid change (marked in green) served as a donor template for HDR. (**B**) Schematic representation of procedures used to generate HEK 293T cells with biallelic mutation at the *HTT locus*. (**C**) RT-PCR analysis of the edited *HTT* gene fragment containing 41, 53 or 84 CAG repeats. The PCR product comprises CAG repeat tract. Unmodified HEK 293T cells contain 16/17 CAG repeats at the *HTT locus*. (**D**) RT-qPCR analysis of the *HTT* mRNA level in 41 CAG, 53 CAG and 84 CAG cell lines. PCR primers were located downstream (3' CAG) from the CAG repeats. All samples were normalized to β-actin, and the results are the mean (± SEM) relative to unmodified HEK 293T cells ($n = 3$; one-way ANOVA followed by Bonferroni's post hoc test; * $p = 0.01$ to 0.05, ** $p = 0.001$ to 0.01, *** $p < 0.001$). (**E**) Western blot analysis of the huntingtin level in the edited cell lines and (**F**) cells transfected with the siRNA_A2 and siHTT. Plectin was used as a loading control. The results indicate the mean (± SEM) relative to cells treated with the control siRNA (BlockIT) ($n = 3$; one way ANOVA followed by Bonferroni's post hoc test; *** $p < 0.001$). (**G**) Aberrant splicing of *HTT* mRNA results in appearance of abnormal transcript containing early intron 1. Analysis of the early intron 1 transcript by RT-PCR assays with three primer pairs spanning the exon 1–exon 2 junction (-17f/exon2r), the exon 1–intron (Fsp2/Rsp2) and intron 1 sequences (2805f/2959r). β-actin was used as a loading control.

2.2. Correction of Mutation in HD Patient-Derived hiPSCs

To correct mutation in HD hiPSCs and generate isogenic control lines, we used CRISPR-Cas9 technology and a well-characterized patient-derived iPSC clonal line (ND42222, 19/109 CAG repeats in *HTT*). Because homologous recombination (HR) occurs with low efficiency, we used various genome-editing approaches and tools to induce DSBs and HDR. These strategies are summarized in Table S1. The first strategy involved a pair of sgRNAs (HTT_sg1 and HTT_sg4) and Cas9n expressed from a plasmid (described also in [32]). As a donor template, we used a 180 nt single-stranded oligodeoxynucleotide (ssODN) containing 10 CAG repeats. The electroporation efficiency was very low, and we observed cell death. This strategy resulted in a lack of correction by HDR, as we identified cell clones with CAG tract excision and perfect strand rejoining. The modification of the ssODN donor template by the extension of homologous arms and the introduction of a silent mutation in a PAM to avoid the recutting of the corrected alleles did not result in genetically corrected clones. In the next strategy, we used RNP complexes composed of the Cas9 protein and synthetic sgRNA targeting the *HTT* sequence upstream of the CAG tract (HTT_sg3). As a donor template, we used a shorter variant (180 nt) of ssODN. At 4 h before electroporation, cells were treated with nocodazole, which arrests S/M phases of the cell cycle and increases the frequency of HDR events. Among the 163 colonies screened,

five were modified by HDR and contained 10/19 CAG, 19/19 CAG or 109/109 CAG repeats in *HTT*, but all of them except for the 19/19 CAG colony contained indel mutations at the cut site. HD hiPSCs were also electroporated with the Cas9 RNP and the HR donor plasmid carrying exon 1 of the *HTT* gene containing 19 CAG repeats (and silent mutation in a PAM). This strategy proved to be the most efficient; among the 131 colonies screened, eight were genetically corrected and contained 19/19 CAG repeats at the *HTT* locus. Three clonal cell lines were selected for more detailed analysis.

The newly generated isogenic control iPSCs were characterized by Sanger sequencing, *HTT* transcript and protein detection, the verification of pluripotency marker expression and karyotype analysis. These lines included iPSCs containing 19/19 CAG repeats (HD_19/19; clones C31.9 and C105), a similar cell line with a silent mutation present in one allele (HD_19/19mut; clone C39) and iPSCs with a double knock-out of the *HTT* gene (HD_KO; clone C37), which was achieved *via* the-out-of-frame excision of the CAG repeat tract with Cas9n, HTT_sg1 and HTT_sg4. Sequencing confirmed the purity of the clones, the repeat tract length and the presence of the silent mutation in C39 clone (Figure S3). This mutation results from homologous recombination with a donor plasmid carrying the silent mutation in a PAM. By western blotting, we confirmed either the presence of normal huntingtin in HD_19/19 (C31.9 and C105 clones) and HD_19/19mut (C39 clone) iPSCs or the absence of huntingtin in HD_KO iPSCs (C37 clone) (Figure 2A). Expression of *HTT* transcript was confirmed by RT-qPCR (Figure 2B). A lack of the most commonly occurring karyotypic abnormalities was observed in most of generated cell lines, as indicated by qPCR-based analysis (Figure S4). The exception was clone C31.9 with possible amplification of analyzed region at chromosome 4. *NANOG*, *OCT4* and *SOX2* markers showed similar expression level in generated lines, compared to that in parental hiPSCs, as analyzed by RT-qPCR (Figure 2C). Expression of pluripotency markers was also confirmed by positive staining for the nuclear markers OCT4 and NANOG and the surface proteins TRA 1–80 and TRA 1–60 (Figure 2D, Figure S5).

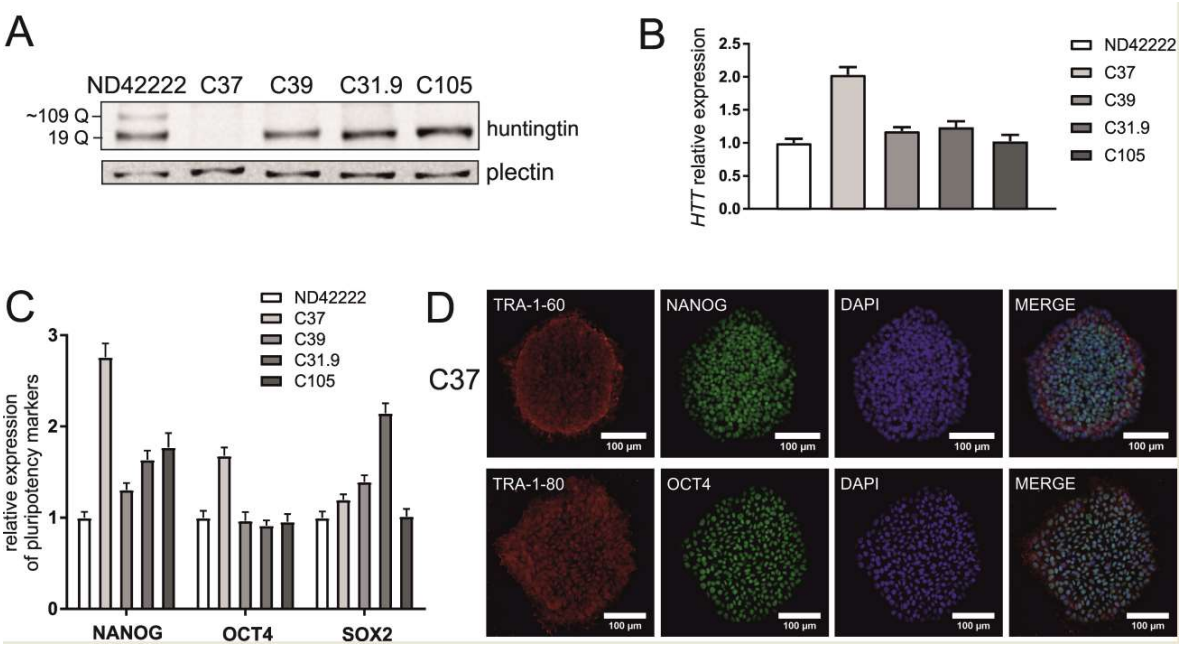


Figure 2. Characteristics of isogenic hiPSC clones. (**A**) Huntingtin protein level in HD hiPSC line with 19/109 Qs and its isogenic control cell lines was analyzed by western blotting. Plectin was used as a reference protein. (**B**) Analysis of the relative expression of the *HTT* gene by RT-qPCR in parental hiPSCs and generated cell lines. (**C**) The gene-edited hiPSCs maintain pluripotency as shown by expression of *NANOG*, *OCT4* and *SOX2* genes and (**D**) positive immunostaining for the pluripotency markers (presented for the C37 clonal cell line). The RT-qPCR results indicate the mean (± SEM) relative to expression level in ND42222, set at 1. ND42222 – parental HD hiPSCs; isogenic cell lines: C37 – HTT_KO hiPSCs, C39 – HD_19/19mut hiPSCs, C31.9 and C105 - HD_19/19 hiPSCs.

To study the potential off-target effects of the CRISPR-Cas9 system, we performed whole-exome sequencing (WES) of three isogenic control hiPSC clones (C37, C39 and C31.9) and a parental patient-derived iPSCs (ND42222) (Table S2). We did not observe mutations resulting from genome editing. A low number of single nucleotide variants (SNVs) were detected in corrected clones (Table S3) with 26 SNVs common to all three cell lines (Table S4).

3. Discussion

Here, we demonstrate the generation of new models of HD based on HEK 293T and hiPSCs. We employed a number of genome-editing approaches and tools, such as wtCas9 and Cas9n, different sgRNAs expressed from plasmids or delivered as RNP complexes, and different HR donor templates in the form of ssODNs or plasmid-based templates. A method involving the use of RNP complexes and an HR donor plasmid with a silent mutation in a PAM sequence was the most efficient in the case of both cell models. This strategy is safe and reduces the possibility of off-target mutations due to the short-term activity of the Cas9 protein and sgRNA delivered in an RNP complex. To induce DSBs, we used HTT_sg3 and the 5'AGG3' PAM sequence located ~20 nt upstream of the CAG tract. In our previous study using the qEva-CRISPR assay, we demonstrated that HTT_sg3 did not induce nonspecific modifications in the tested off-target regions [32]. Other strategies were less efficient in the generation of successful edits; e.g., the use of double nickase strategy resulted in the preferential excision of CAG repeats, strand rejoining and *HTT* knock-out.

In our study, a series of isogenic HEK 293T cell clones containing expanded CAG repeats at the *HTT* locus was generated for the first time. The lengths of the repeat tracts represent frequent variants observed in HD patients (41 and 53 CAG repeats) and in a juvenile form of HD (84 CAG repeats). It is worth noting that each clone has a homozygous genotype in which the two alleles harbor a repeat tract of the same length. This characteristic makes these cell lines very useful considering the technical difficulties encountered in methods related to repeated sequences, as observed in PCR, sequencing and western blotting.

HEK 293 cells present some characteristics of neuronal lineage cells, such as the potential to propagate highly neurotropic viruses and inducible synaptogenesis [33]. It has been demonstrated that edited HEK 293 cells containing ~100 and 150 CAG repeats at the *HTT* locus undergo a wide spectrum of pathological changes characteristic of HD [25]. Despite the fact that HEK 293T cells differ significantly from neurons, which are the main site of HD pathogenesis, we observed the presence of the abnormal transcript resulting from the aberrant splicing of *HTT* mRNA. The same transcripts were detected in patient-derived fibroblasts, postmortem HD brains and mouse models expressing mutant *Htt* (mouse) or *HTT* (human) [31]. Nonetheless, HEK 293T cells, even with mutation in *HTT* gene, are not a good model to study some aspects of HD pathogenesis. The set of generated models will be valuable rather to study CAG repeat expansion/contraction mechanisms, aberrant splicing of *HTT* transcript, RAN translation, frameshifting or to test various huntingtin lowering therapeutic strategies. More appropriate models to study HD pathogenesis are neural cells derived from isogenic hiPSCs containing the same genetic background.

Using a similar strategy involving the CRISPR-Cas9 system delivered in an RNP complex and HR donor plasmid, we corrected HD iPSCs and generated healthy isogenic controls. In addition, by using an sgRNA pair and Cas9n, we excised the CAG repeat tract and generated a double knock-out of the *HTT* gene in parental HD iPSCs. HDR was inefficient under this approach. In a previous study involving the generation of isogenic HD control hiPSCs [21], the authors used Cas9n, a pair of sgRNAs and a piggyBac (PB) transposon selection cassette-based HR donor template. After the puromycin selection of edited clones and the excision of the selection cassette by transient transfection with a PB-expressing plasmid, 4.7% of clones were positively verified by immunoblotting. In our study, we used a much simpler approach involving the Cas9/sgRNA RNP complex without the need for additional selection steps, achieving a similar efficiency of editing (6%). We confirmed that the modified isogenic hiPSC clones retain pluripotency and a normal karyotype, and we demonstrated

the expression of normal *HTT* in corrected control hiPSCs. Whole-exome sequencing demonstrated variability mainly due to clonal differences and the method imperfections, as we did not identify potential off-target sites in any of the lists of detected variants.

The parental HD hiPSC line ND42222 was recently characterized in detail using transcriptomics and proteomics approaches [34]. A number of deregulated genes were identified compared to healthy control hiPSCs. It would be interesting to validate these data by using isogenic controls to exclude the effects of the genetic background. Moreover, after the neural differentiation of hiPSCs, a set of isogenic cell lines, including a mutant line, a normal line and an *HTT* knock-out line, will be a valuable model for studying various aspects of HD.

4. Materials and Methods

4.1. HEK 293T Cell Culture And siRNA Transfection

HEK 293T cells containing 16/17 CAG repeats in the *HTT* gene (ATCC, Manassas, VA, USA) were grown in Dulbecco's modified Eagle's medium (Lonza; Basel, Switzerland) supplemented with 10% fetal bovine serum (FBS) (Sigma-Aldrich, St. Louis, MO, USA), antibiotics (Sigma-Aldrich) and L-glutamine (Sigma-Aldrich). All RNA oligonucleotides were synthesized at Future Synthesis (Poznan, Poland). Briefly, RNAs were combined in annealing buffer (Thermo Fisher Scientific, Waltham, MA, USA) to a 20-µM duplex concentration and incubated at 90 °C for 1 min, followed by additional incubation at room temperature for 45 min. The sequences of the siRNAs used in this study are presented in [27,28]. At 24 h prior to transfection 3×10^5 cells were seeded on 6 cm plates. Cells transfections were performed using 100 nM siRNAs and Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's instructions. The transfection efficiency was monitored using 20 nM BlockIT fluorescent siRNA (Life Technologies, Carlsbad, CA, USA). Due to the rapid growth of the 41 CAG, 53 CAG and 84 CAG cell lines, the medium was changed after 4 h from transfection for the complete medium containing 4% FBS. The efficiency of silencing was analyzed 48 h post transfection by western blotting.

4.2. Donor Template

Single-stranded oligodeoxynucleotides (ssODNs) were synthesized (IDT, Skokie, IL, USA). The HR donor plasmid was prepared by cloning the PCR products with 41 and 85 CAG repeats (for HEK 293T cells) and 19 CAG repeats (for hiPSCs) and asymmetric homologous arms (139 bp and 375 bp) into the pGEM-T easy vector (Promega, Madison, WI, USA). During the transformation of GT116 *E. coli* cells, 85 CAG repeats were shortened to 53 and 84 CAG repeats. Finally, plasmids with 41, 53 and 84 CAG repeats (for HEK 293T) or 19 CAG repeats (for hiPSCs) were digested with the SacII enzyme (Thermo Fisher Scientific) and used as donor templates for further experiments. The PAM sequence in the donor template was mutated (5'AGG3' to 5'ACG3') using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent, Santa Clara, CA, USA) and mutHDg3F/mutHDg3R primers, to avoid nonspecific cutting of the plasmid by CRISPR-Cas9. To increase the frequency of HDR, cells were synchronized and arrested in G2/M phase by using 40 nM nocodazole (Sigma-Aldrich) at 4 h before electroporation.

4.3. HTT Gene Editing with the Plasmid–Based CRISPR-Cas9 System

The guide RNAs specific for the *HTT* gene (HTT_sg1, HTT_sg3 and HTT_sg4) were previously described and validated [26,32,35]. The top and bottom strands of the 20-nt guide RNAs (Table S5) were synthesized (IBB, Warsaw, Poland), annealed and ligated into the pair of FastDigest BpiI (Thermo Fisher Scientific) cut plasmids: pSpCas9(BB)-2A-GFP (PX458) and its nickase version (D10A nickase mutant; pSpCas9n(BB)-2A-GFP (PX461)) (Addgene, Cambridge, MA, USA) from *S. pyogenes* [36]. The ligated products were transformed into chemically competent *E. coli* GT116 cells (InvivoGen, San Diego, CA, USA), and the cells were plated onto ampicillin selection plates (100 µg/mL ampicillin) and incubated at 37°C overnight. Plasmid DNA was isolated using the Gene JET Plasmid Miniprep kit (Thermo Fisher

Scientific) and verified with Sanger sequencing. Electroporation was used to deliver Cas9 protein, HTT_sgRNA and a donor template for HDR.

4.4. Editing of HEK 293T Cells with an RNP Complex

Cells were electroporated with an RNP complex composed of SpCas9, crRNA (HTT_sg3) and fluorescent tracer RNA, ATTO550, which is a novel fluorescent label related to the well-known dyes rhodamine 6G and rhodamine B (IDT), with 600 ng of a linearized HR donor plasmid. Before electroporation, CRISPR RNA (crRNA) and trans-activating small RNA (tracrRNA) oligos were mixed at an equimolar ratio in nuclease-free duplex buffer (IDT) to achieve a final concentration of the gRNA complex of 60 µM. The crRNA and tracrRNA duplex was heated at 95 °C for 5 min following 10 min of incubation at room temperature. The RNP complex was produced by mixing 5 µg (~30 pmol) of the recombinant NLS-SpCas9-NLS nuclease (VBCF Protein Technologies facility http://www.vbcf.ac.at) and 60 pmol of sgRNA, followed by incubation at room temperature for 10-20 min. HEK 293T cells were electroporated with the NeonTM Transfection System (Invitrogen, Carlsbad, CA, USA). Briefly, 2×10^5 cells were harvested, resuspended in buffer R and electroporated with the RNP complex and 600 ng of the donor template in 10 µL tips using the following parameters: 1.150 V, 20 ms, 2 pulses. After electroporation, the cells were seeded at a low density (1–2 × 10^3 cells/10 cm plate), and after 2 h, attached cells were identified based on the presence of a red signal (from fluorescent tracrRNA) during microscopic observation under a UV lamp. Monoclonal culturing was carried out for approximately 1.5 weeks, after which colonies were transferred to 48-well plates. The monoclonal cultures were analyzed after reaching approximately 60−80% confluency.

4.5. Generation of Human iPS Cells with Modifications in the HTT Gene

Parental HD iPSCs (ND42222) obtained from the NINDS Human Genetic Resource Center (Coriell Institute, Camden, NJ, USA) were characterized by NINDS via the analysis of pluripotency marker expression, colony formation, and karyotyping. Cells were grown in Essential 8 medium (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) on Geltrex-coated dishes (Gibco). At 24 h before electroporation, after reaching 70% confluence, the medium was replaced with Essential 8 medium supplemented with the ROCK pathway inhibitor Y-27632 (STEMCELL Technologies, Vancouver, Canada) (final concentration 10 µM) to increase the survival of cells after electroporation. At 4 h before electroporation, the medium was supplemented with nocodazole (40 nM) [37,38]. Then, the cells were dissociated to single cells by incubation with 0.5 mM PBS-EDTA solution for 10 min. Next, 2×10^5 cells were harvested, resuspended in buffer R and electroporated with the RNP complex and 600 ng of the donor template. Electroporation was conducted in 10 µL tips using the following parameters: 1100 V, 20 ms, and 2 pulses. After electroporation, the cells were seeded on Geltrex-coated dishes at low densities (2–15 × 10^3 cells/10 cm plate). Four hours later, after the attachment of the cells, the medium was replaced with StemFlex Medium (Gibco) containing Y-27632. Then, cells exhibiting red signals were identified. The medium was replaced every 3–4 days until newly formed colonies exhibited sufficient growth for transfer. Monoclonal cultures derived from single iPSCs were manually picked and transferred to Geltrex-coated 48-well plates. Monoclonal cultures were grown in StemFlex Medium until they reached 80% confluence and then analyzed.

4.6. DNA Extraction and PCR

Genomic DNA from the HEK 293T and iPSC monoclones was extracted using QuickExtract™ DNA Extraction Solution (Lucigen, Middleton, WI, USA) according to the manufacturer's instructions. Genomic DNA was amplified using GoTaq® G2 DNA Polymerase (Promega) with primers HD1F and HD1R spanning the CAG repeats in exon 1 of the *HTT* gene. The PCR amplification program was as follows: initial denaturation at 95 °C for 3 min; 30 cycles at 95 °C for 30 s, 62 °C for 30 s, and 72 °C for 45 s; and a final elongation at 72 °C for 5 min. The same conditions were used for RT-PCR, with annealing temperature 59 °C for all primer pairs (−17f and Exon2r, 2805f and 2959r and Fsp2 and Rsp2). The

PCR products were separated in 1.3% agarose gels and detected using UV transilluminator G:BOX (Syngene, Cambridge, UK). The PCR products of selected clones were purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific) and sequenced with the HD1F primer. The primer sequences are listed in the Table S6.

4.7. RNA Extraction and RT-qPCR

Total RNA was isolated from HEK 293T cells using the TRI Reagent (BioShop, Burlington, Canada) according to the manufacturer's instructions. The RNA concentration was measured using a spectrophotometer (DeNovix, Wilmington, NC, USA). A total of 700 ng (HEK 293T) or 500 ng of RNA (hiPSCs) was reverse transcribed at 55 °C using Superscript III (Life Technologies) and random hexamer primers (Promega). The quality of the reverse transcription (RT) reaction was assessed through polymerase chain reaction (PCR) amplification of the β-actin gene. Complementary DNA (cDNA) was employed for quantitative polymerase chain reaction (qPCR) using SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad, Hercules, CA, USA) with denaturation at 95 °C for 30 s, followed by 40 cycles of denaturation at 95 °C for 15 s and annealing at 60 °C for 30 s. The melt curve protocol was subsequently performed with *HTT*-, pluripotency markers-, or *β-actin* or *GAPDH*-specific primers as follows: 5 s at 65 °C, followed by 5 s increments at 0.5 °C from 65 °C to 95 °C, in the CFX Connect™ Real-Time PCR Detection System (Bio-Rad). The primers used for RT-qPCR were designed to cover the *HTT* region downstream (HD 3'CAG) of the CAG repeat tract. The sequences of the primers are presented in Supplementary Table S6. Data preprocessing and normalization were performed using Bio-Rad CFX Manager software (Bio-Rad).

4.8. Immunocytochemistry

Generated iPSCs were plated on Matrigel (Corning, NY, USA)-coated cover slips and grown in Essential 8 medium. Then, the cells were fixed in 4% PFA, permeabilized with 0.5% Tween and blocked in 1% bovine serum albumin. Next, the cells were incubated overnight with the primary antibodies (listed in Table S7) at 4 °C. Thereafter, the cells were washed 3× with PBS for at least 5 min each time and incubated 1 h with fluorescent-dye conjugated secondary antibodies at room temperature and again washed 3× with PBS for at least 5 min. SlowFade Diamond Antifade Mountant with DAPI (Invitrogen) was used for nuclear staining. Images were captured on a Leica SP5 confocal microscope.

4.9. Western Blotting

Protein was isolated from the cells with the use of PB (60 mM Tris-base, 2% SDS, 10% sucrose, 2 mM PMSF). A total of 30 μg of protein was resolved on a Tris-acetate sodium dodecyl sulfate (SDS)-polyacrylamide gel (1.5 cm, 4% stacking gel/4.5 cm, 5% resolving gel, acrylamide:bis-acrylamide ratio of 49:1) in XT Tricine buffer (Bio-Rad) at 135 V in an ice-water bath. After electrophoresis, the proteins were wet-transferred overnight to a nitrocellulose blotting membrane (GE Healthcare Life Sciences, Chicago, Illinois, IL, USA). The primary antibodies, including anti-huntingtin, anti-plectin, and the anti-rabbit HRP conjugate secondary antibody were used in a PBS/0.1% Tween-20 buffer containing 5% nonfat milk. The immunoreaction was detected using the Westar Antares Chemiluminescent substrate (Cyanagen, Bologna, Italy). The protein bands were scanned directly from the membrane using a camera and quantified using a Gel-Pro Analyzer (Media Cybernetics, Rockville, MD, USA). A list of all antibodies used is provided in Table S7.

4.10. Karyotyping

Genomic DNA was isolated with a Genomic DNA Isolation Kit (Norgen Biotek, Schmon Parkway, Thorold, Canada). Then, a qPCR-based hPSC Genetic Analysis Kit (STEMCELL Technologies) was used to detect the major potential karyotypic abnormalities reported in human iPSCs according to the manufacturer's instructions. Because the designed genome editing procedure was performed within

chromosome 4, we used the Chr1q region as an internal control instead of the default region of Chr4p in calculations concerning potential abnormalities.

4.11. Whole-Exome Sequencing and Data Analysis

WES analysis was performed by CeGaT GmbH (Tubingen, Germany). 50 ng of high-molecular weight DNA per sample were used for preparing exome-enriched libraries with Twist Human Core Exome kit (Twist Bioscience, San Francisco, CA, USA). This kit is designed to target 33 Mb of highly conserved protein-coding regions. DNA fragmentation was performed using an enzymatic reaction. Subsequently, end-repair, dA-tailing, index adaptor ligation and purification was performed. These steps are followed by pre-capture PCR amplification and bead-based purification. The library was quantified using a Qubit dsDNA Broad Range Quantitation Assay and an average fragment length of 375 bp to 425 bp was ensured. Amplified, indexed libraries were pooled and hybridized to capture probes and bound to streptavidin binding beads. As a next step, post-capture PCR amplification was performed. After purification, the libraries were quantified using the Agilent BioAnalyzer High Sensitivity DNA Kit and a Thermo Fisher Scientific Qubit dsDNA High Sensitivity Quantitation Assay. Sequencing was performed on a NovaSeq 6000 (Illumina, San Diego, CA, USA) and 14.8 Gb (C31.9), 9.5 Gb (C39), 10.5 Gb (ND42222), and 12.2 Gb (C37) of data were produced. This resulted in an average coverage of 134, 91.7, 110.6, and 114.9, respectively.

Trimmed raw reads were aligned to the human reference genome (hg19-cegat) using the Burrows-Wheeler Aligner (BWA-mem version 0.7.17-cegat) [39]. ABRA (version 2.18) [40] was used for local realignment of reads in target regions to facilitate more accurate indel calling. In the reference hg19-cegat the pseudo-autosomal regions (PAR) on chromosome Y were masked (chrY:10001-2649520, chrY:59034050- 59363566). This procedure prevents reads that map to this region from being discarded due to mapping to two different chromosomes. Reads that could be aligned to more than one locus with the same mapping score were discarded. Duplicated reads, which most likely originated from the same PCR amplicon, were discarded as well. A proprietary software was used for variant detection (observed frequency of the alternative allele (OFA) > 0.85). Variants that occurred in the control sample ND42222 were excluded from the lists of the treated samples.

HTT_sg1, HTT_sg3 and HTT_sg4 were used to conduct an off-target site search using CCTop [41] allowing for up to four mismatches. Tables containing insertions and deletions as well as SNVs and the lists of potential off-target sites were imported into R version 3.6.1. Positions were then compared using the dplyr package. No position that was identified as a potential off-target site was present in any of the lists.

4.12. Mycoplasma Testing

Cultures were tested for mycoplasma contamination using a Veno GeM Classic Mycoplasma PCR detection Kit (Minerva Biolabs, Berlin, Germany) according to the manufacturer's instructions.

4.13. Statistical Analysis

Statistical analysis was performed using GraphPad Prism v. 5.0 software (GraphPad, San Diego, CA, USA). Data were analyzed using one-way ANOVA followed by Bonferroni's post hoc test (p-value: ns >0.05, * 0.01 to 0.05, ** 0.001 to 0.01, *** $p < 0.001$) with an arbitrary value of 1 assigned to the unmodified HEK 293T cells (Figure 1D) and to the cells treated with control (non-targeting) siRNA (Figure 1F).

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/21/5/1854/s1. Figure S1. Sanger sequencing analysis of the HTT locus in modified HEK 293T cell lines (41 CAG, 53 CAG and 84 CAG); Figure S2. Western blot analysis of HTT protein downregulation in edited HEK 293T cells treated with siRNA_A2 (A2) and siHTT. C – cells treated with control siRNA (without target). Plectin was used as a loading control; Figure S3. Sanger sequencing analysis of edited hiPSC lines; Figure S4. Results from qPCR-based analysis of potential karyotypic abnormalities in generated isogenic hiPSC lines. In case of C31.9 clone possible

amplification of analyzed region at chromosome 4 is observed; Figure S5. The gene-edited hiPSCs maintain pluripotency as shown by positive immunostaining for the pluripotency markers; Table S1. Editing strategies used to generate isogenic models of HD in HEK 293T cells and hiPSCs; Table S2. Summary of capture statistics for whole-exome sequencing; Table S3. Sequence Variants in the gene-corrected hiPSC clones by whole-exome sequence analysis; Table S4. Sequence variants present in all edited clones C37, C39, C31.9 detected by whole-exome sequencing; Table S5. DNA and RNA oligonucleotides used for the generation of sgRNAs; Table S6. DNA oligonucleotides used as primers for PCR, RT-qPCR and directed mutagenesis; Table S7. Antibodies used for immunocytochemistry (ICC) and western blotting (WB).

Author Contributions: M.O., A.F. and M.D. conceived and designed the study. M.D., A.C. and E.K. performed the experiments. M.D., A.C., M.O. and A.F. analyzed the data. M.O. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by research grants from the National Science Center PL (2015/18/E/NZ2/00678; 2018/29/B/NZ1/00293; 2015/17/D/NZ5/03443, 2015/17/N/NZ2/01916 and 2015/19/B/NZ2/02453); Funding for open access charge: National Science Center (2018/29/B/NZ1/00293).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Macdonald, M.E.; Ambrose, C.M.; Duyao, M.P.; Myers, R.H.; Lin, C.; Srinidhi, L.; Barnes, G.; Taylor, S.A.; James, M.; Groot, N.; et al. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell* **1993**, *72*, 971–983. [CrossRef]
2. Rubinsztein, D.C.; Leggo, J.; Coles, R.; Almqvist, E.; Biancalana, V.; Cassiman, J.-J.; Chotai, K.; Connarty, M.; Craufurd, D.; Curtis, A.; et al. Phenotypic characterization of individuals with 30–40 CAG repeats in the huntington disease (HD) gene reveals HD cases with 36 repeats and apparently normal elderly individuals with 36–39 repeats. *Am. J. Hum. Genet.* **1996**, *59*, 16–22. [PubMed]
3. Kremer, B.; Goldberg, P.; Andrew, S.E.; Theilmann, J.; Telenius, H.; Zeisler, J.; Squitieri, F.; Lin, B.; Bassett, A.; Almqvist, E.; et al. A worldwide study of the huntington's disease mutation: The sensitivity and specificity of measuring CAG repeats. *N. Engl. J. Med.* **1994**, *330*, 1401–1406. [CrossRef] [PubMed]
4. Davies, S.W.; Turmaine, M.; Cozens, B.A.; DiFiglia, M.; Sharp, A.H.; Ross, C.A.; Scherzinger, E.; Wanker, E.; Mangiarini, L.; Bates, G. Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell* **1997**, *90*, 537–548. [CrossRef]
5. Scherzinger, E.; Lurz, R.; Turmaine, M.; Mangiarini, L.; Hollenbach, B.; Hasenbank, R.; Bates, G.; Davies, S.W.; Lehrach, H.; Wanker, E. Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. *Cell* **1997**, *90*, 549–558. [CrossRef]
6. Ross, C.A. Polyglutamine pathogenesis: Emergence of unifying mechanisms for Huntington's disease and related disorders. *Neuron* **2002**, *35*, 819–822. [CrossRef]
7. Zhang, N.; Bailus, B.J.; Ring, K.L.; Ellerby, L.M. IPSC-based drug screening for Huntington's disease. *Brain Res.* **2015**, *1638*, 42–56. [CrossRef]
8. Consortium GM of HD; Lee, J.-M.; Correia, K.; Loupe, J.; Kim, K.-H.; Barker, D.; Hong, E.; Chao, M.; Long, J.; Lucente, D.; et al. Huntington's disease onset is determined by length of uninterrupted CAG, not encoded polyglutamine, and is modified by DNA maintenance mechanisms. *bioRxiv* **2019**. [CrossRef]
9. Wild, E.J.; Tabrizi, S.J. Therapies targeting DNA and RNA in Huntington's disease. *Lancet Neurol.* **2017**, *16*, 837–847. [CrossRef]
10. Tabrizi, S.J.; Ghosh, R.; Leavitt, B.R. Huntingtin Lowering Strategies for Disease Modification in Huntington's Disease. *Neuron* **2019**, *101*, 801–819. [CrossRef]
11. Byrne, S.; Mali, P.; Church, G.M. Genome Editing in Human Stem Cells. *Methods Enzymol.* **2014**, *546*, 119–138. [CrossRef] [PubMed]
12. Dastidar, S.; Ardui, S.; Singh, K.; Majumdar, D.; Nair, N.; Fu, Y.; Reyon, D.; Samara, E.; Gerli, M.F.M.; Klein, A.; et al. Efficient CRISPR/Cas9-mediated editing of trinucleotide repeat expansion in myotonic dystrophy patient-derived iPS and myogenic cells. *Nucleic Acids Res.* **2018**, *46*, 8275–8298. [CrossRef]
13. Park, C.-Y.; Halevy, T.; Lee, D.R.; Sung, J.J.; Lee, J.S.; Yanuka, O.; Benvenisty, N.; Kim, D.-W. Reversion of FMR1 Methylation and Silencing by Editing the Triplet Repeats in Fragile X iPSC-Derived Neurons. *Cell Rep.* **2015**, *13*, 234–241. [CrossRef] [PubMed]

14. Raaijmakers, R.H.L.; Ripken, L.; Ausems, R.; Wansink, D.G. CRISPR/Cas Applications in Myotonic Dystrophy: Expanding Opportunities. *Int. J. Mol. Sci.* **2019**, *20*, 3689. [CrossRef] [PubMed]
15. Marthaler, A.G.; Tubsuwan, A.; Schmid, B.; Poulsen, U.B.; Engelbrecht, A.F.; Mau-Holzmann, U.A.; Hyttel, P.; Nielsen, T.T.; Nielsen, J.E.; Holst, B. Generation of an isogenic, gene-corrected control cell line of the spinocerebellar ataxia type 2 patient-derived iPSC line H266. *Stem Cell Res.* **2016**, *16*, 202–205. [CrossRef] [PubMed]
16. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J.A.; Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337*, 816–821. [CrossRef]
17. Takahashi, K.; Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **2006**, *126*, 663–676. [CrossRef]
18. Wu, Y.-Y.; Chiu, F.-L.; Yeh, C.-S.; Kuo, H.-C. Opportunities and challenges for the use of induced pluripotent stem cells in modelling neurodegenerative disease. *Open Boil.* **2019**, *9*, 180177. [CrossRef]
19. Jaworska, E.; Kozlowska, E.; Switonski, P.M.; Krzyzosiak, W.J. Modeling simple repeat expansion diseases with iPSC technology. *Cell. Mol. Life Sci.* **2016**, *73*, 4085–4100. [CrossRef]
20. An, M.C.; Zhang, N.; Scott, G.; Montoro, D.; Wittkop, T.; Mooney, S.; Melov, S.; Ellerby, L.M. Genetic correction of Huntington's disease phenotypes in induced pluripotent stem cells. *Cell Stem Cell* **2012**, *11*, 253–263. [CrossRef]
21. Xu, X.; Tay, Y.; Sim, B.; Yoon, S.-I.; Huang, Y.; Ooi, J.; Utami, K.H.; Ziaei, A.; Ng, B.; Radulescu, C.; et al. Reversal of Phenotypic Abnormalities by CRISPR/Cas9-Mediated Gene Correction in Huntington Disease Patient-Derived Induced Pluripotent Stem Cells. *Stem Cell Rep.* **2017**, *8*, 619–633. [CrossRef] [PubMed]
22. An, M.C.; O'Brien, R.N.; Zhang, N.; Patra, B.N.; De La Cruz, M.; Ray, A.; Ellerby, L.M. Polyglutamine Disease Modeling: Epitope Based Screen for Homologous Recombination using CRISPR/Cas9 System. *PLoS Curr.* **2014**, *6*. [CrossRef] [PubMed]
23. Ooi, J.; Langley, S.R.; Xu, X.; Utami, K.H.; Sim, B.; Huang, Y.; Harmston, N.; Tay, Y.L.; Ziaei, A.; Zeng, R.; et al. Unbiased Profiling of Isogenic Huntington Disease hPSC-Derived CNS and Peripheral Cells Reveals Strong Cell-Type Specificity of CAG Length Effects. *Cell Rep.* **2019**, *26*, 2494–2508. [CrossRef] [PubMed]
24. Olejniczak, M.; Urbanek-Trzeciak, M.; Jaworska, E.; Witucki, L.; Szcześniak, M.; Makalowska, I.; Krzyzosiak, W.J. Sequence-non-specific effects generated by various types of RNA interference triggers. *Biochim. Biophys. Acta Gene Regul. Mech.* **2016**, *1859*, 306–314. [CrossRef]
25. Morozova, K.; Suldina, L.A.; Malankhanova, T.B.; Grigor'Eva, E.V.; Zakian, S.M.; Kiseleva, E.; Malakhova, A.A. Introducing an expanded CAG tract into the huntingtin gene causes a wide spectrum of ultrastructural defects in cultured human cells. *PLoS ONE* **2018**, *13*, e0204735. [CrossRef]
26. Dąbrowska, M.; Juzwa, W.; Krzyzosiak, W.J.; Olejniczak, M. Precise Excision of the CAG Tract from the Huntingtin Gene by Cas9 Nickases. *Front. Mol. Neurosci.* **2018**, *12*. [CrossRef]
27. Wang, Y.-L.; Liu, W.; Wada, E.; Murata, M.; Wada, K.; Kanazawa, I. Clinico-pathological rescue of a model mouse of Huntington's disease by siRNA. *Neurosci. Res.* **2005**, *53*, 241–249. [CrossRef]
28. Fiszer, A.; Olejniczak, M.; Galka-Marciniak, P.; Mykowska, A.; Krzyzosiak, W.J. Self-duplexing CUG repeats selectively inhibit mutant huntingtin expression. *Nucleic Acids Res.* **2013**, *41*, 10426–10437. [CrossRef]
29. Saudou, F.; Humbert, S. The Biology of Huntingtin. *Neuron* **2016**, *89*, 910–926. [CrossRef]
30. Neueder, A.; Dumas, A.A.; Benjamin, A.; Bates, G. Regulatory mechanisms of incomplete huntingtin mRNA splicing. *Nat. Commun.* **2018**, *9*, 3955. [CrossRef]
31. Sathasivam, K.; Neueder, A.; Gipson, T.A.; Landles, C.; Benjamin, A.; Bondulich, M.K.; Smith, N.L.; Faull, R.; Roos, R.A.C.; Howland, D.; et al. Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2366–2370. [CrossRef] [PubMed]
32. Dąbrowska, M.; Czubak, K.; Juzwa, W.; Krzyzosiak, W.J.; Olejniczak, M.; Kozlowski, P. qEva-CRISPR: A method for quantitative evaluation of CRISPR/Cas-mediated genome editing in target and off-target sites. *Nucleic Acids Res.* **2018**, *46*, e101. [CrossRef] [PubMed]
33. Stepanenko, A.; Dmitrenko, V. HEK293 in cell biology and cancer research: Phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution. *Gene* **2015**, *569*, 182–190. [CrossRef] [PubMed]

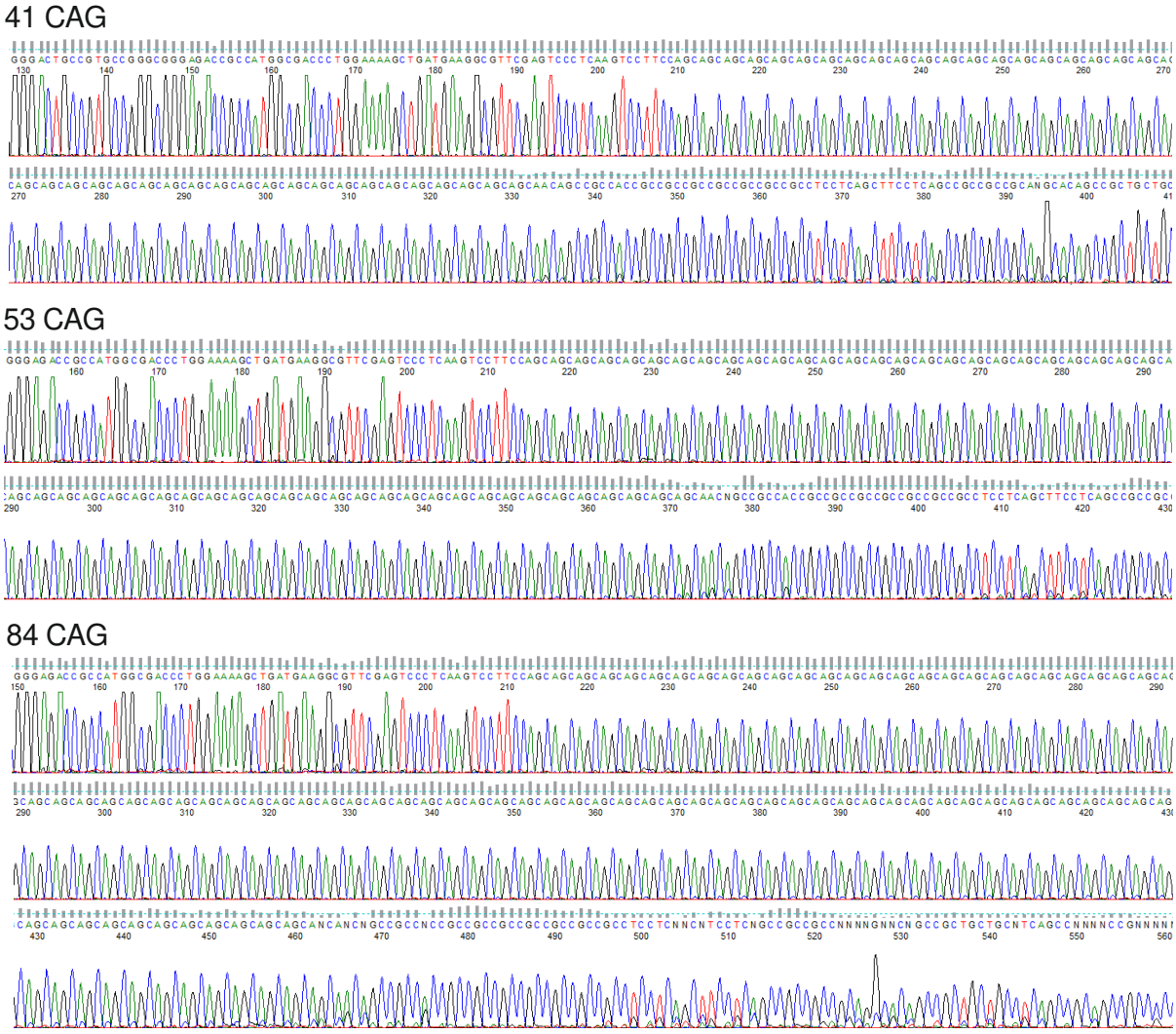
34. Świtońska, K.; Szlachcic, W.J.; Handschuh, L.; Wojciechowski, P.; Marczak, Ł.; Stelmaszczuk, M.; Figlerowicz, M.; Figiel, M. Identification of Altered Developmental Pathways in Human Juvenile HD iPSC With 71Q and 109Q Using Transcriptome Profiling. *Front. Cell. Neurosci.* **2019**, *12*. [CrossRef] [PubMed]
35. Dabrowska, M.; Olejniczak, M. Gene Therapy for Huntington's Disease Using Targeted Endonucleases. *Methods Mol. Biol.* **2020**, *2056*, 269–284. [CrossRef] [PubMed]
36. Ran, F.A.; Hsu, P.; Lin, C.-Y.; Gootenberg, J.S.; Konermann, S.; Trevino, A.E.; Scott, D.A.; Inoue, A.; Matoba, S.; Zhang, Y.; et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **2013**, *154*, 1380–1389. [CrossRef]
37. Yiangou, L.; Grandy, R.A.; Morell, C.M.; Tomaz, R.A.; Osnato, A.; Kadiwala, J.; Muraro, D.; Garcia-Bernardo, J.; Nakanoh, S.; Bernard, W.; et al. Method to Synchronize Cell Cycle of Human Pluripotent Stem Cells without Affecting Their Fundamental Characteristics. *Stem Cell Rep.* **2018**, *12*, 165–179. [CrossRef]
38. Yang, D.; A Scavuzzo, M.; Chmielowiec, J.; Sharp, R.; Bajic, A.; Borowiak, M. Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases. *Sci. Rep.* **2016**, *6*, 21264. [CrossRef]
39. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef]
40. Mose, L.E.; Wilkerson, M.D.; Hayes, D.N.; Perou, C.M.; Parker, J.S. ABRA: Improved coding indel detection via assembly-based realignment. *Bioinformatics* **2014**, *30*, 2813–2815. [CrossRef]
41. Dobson, L.; Reményi, I.; Tusnády, G.E. CCTOP: A Consensus Constrained TOPology prediction web server. *Nucleic Acids Res.* **2015**, *43*, W408-12. [CrossRef] [PubMed]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

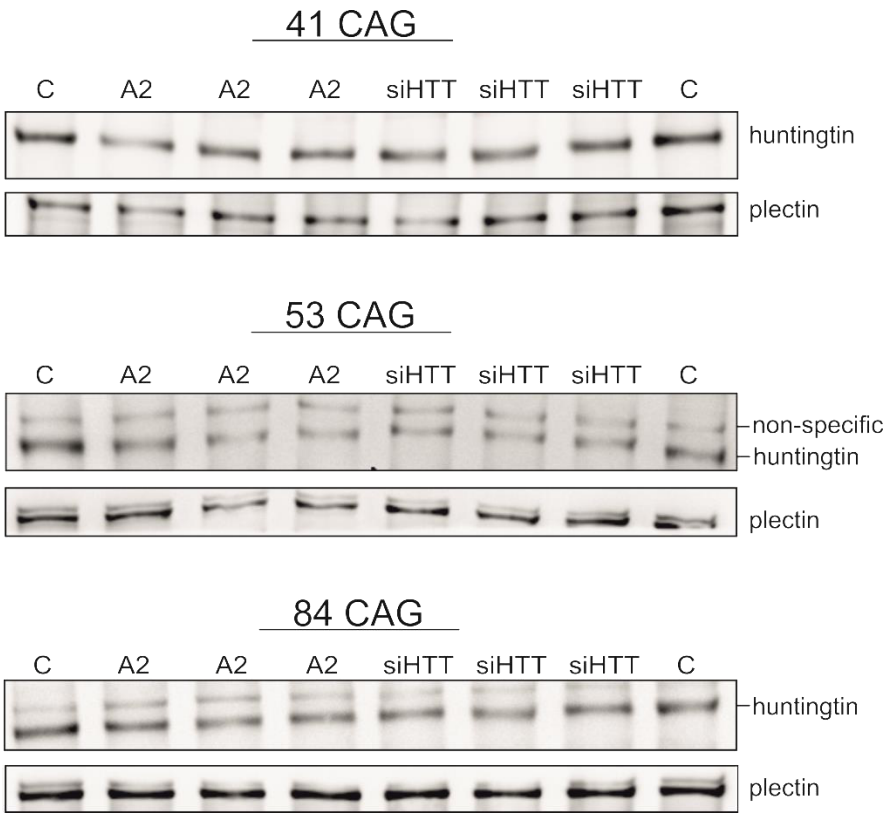
Article

Generation of New Isogenic Models of Huntington's Disease Using CRISPR-Cas9 Technology

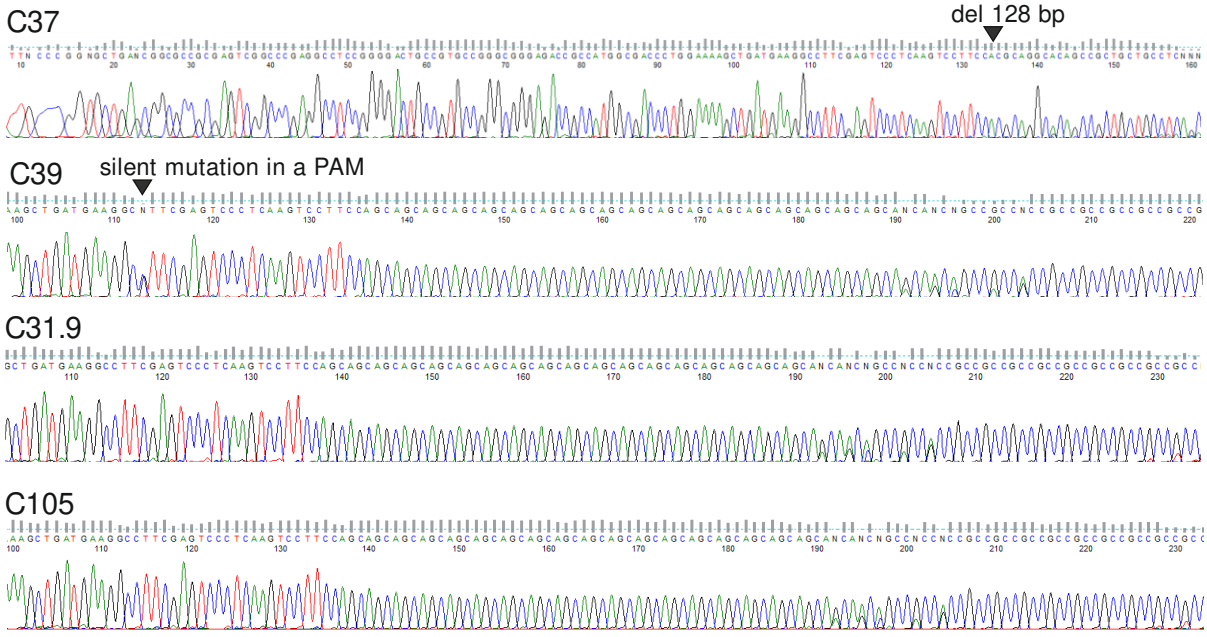
Magdalena Dabrowska [1], Agata Ciolak [2], Emilia Kozlowska [2], Agnieszka Fiszer [2] and Marta Olejniczak [1,*]



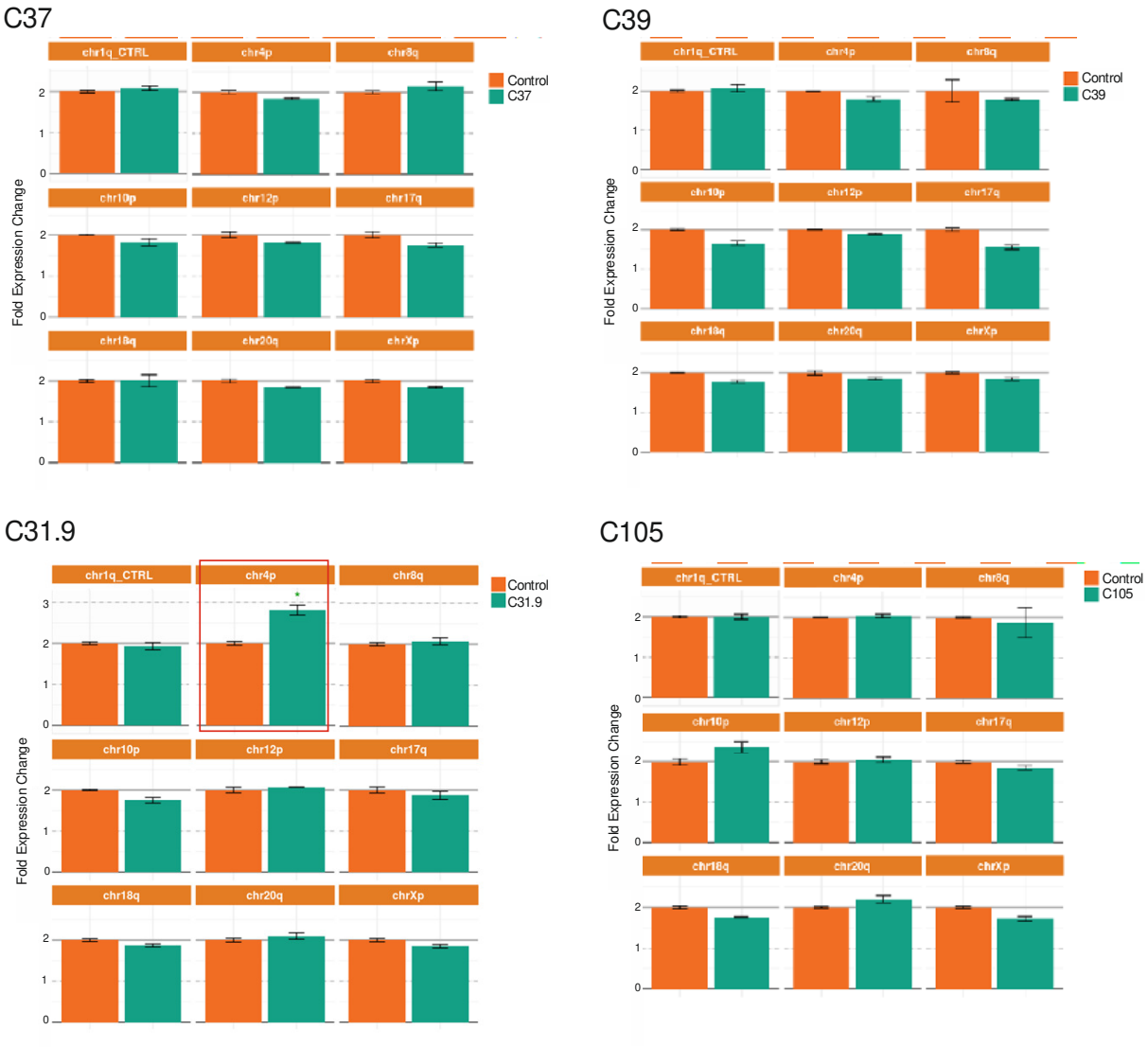
Supplemental Figure S1. Sanger sequencing analysis of the *HTT locus* in modified HEK 293T cell lines (41 CAG, 53 CAG and 84 CAG).



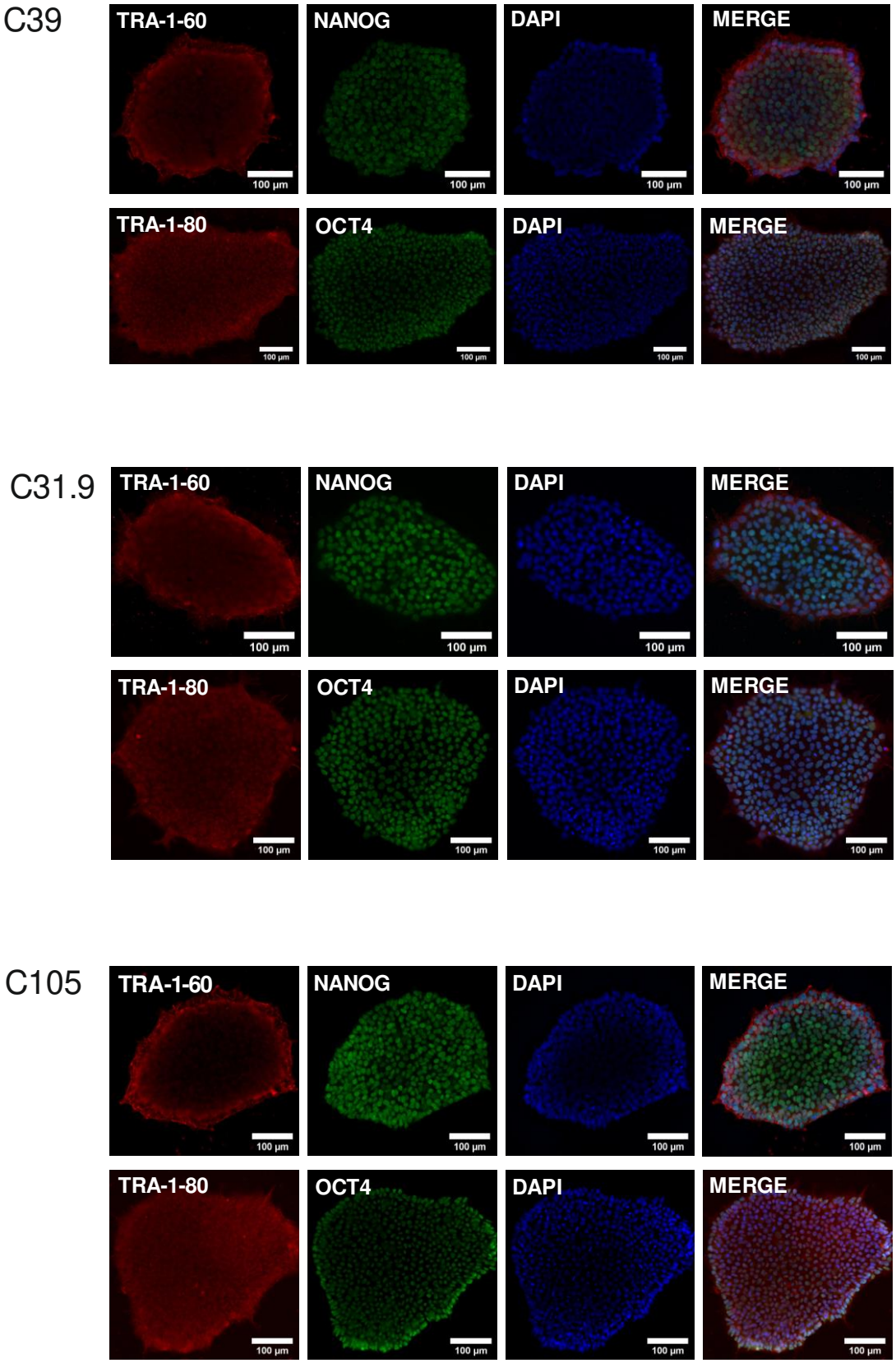
Supplemental Figure S2. Western blot analysis of HTT protein downregulation in edited HEK 293T cells treated with siRNA_A2 (A2) and siHTT. C – cells treated with control siRNA (without target). Plectin was used as a loading control.



Supplemental Figure S3. Sanger sequencing analysis of edited hiPSC lines.



Supplemental Figure S4. Results from qPCR-based analysis of potential karyotypic abnormalities in generated isogenic hiPSC lines. In case of C31.9 clone possible amplification of analyzed region at chromosome 4 is observed.



Supplemental Figure S5. The gene-edited hiPSCs maintain pluripotency as shown by positive immunostaining for the pluripotency markers.

Supplemental Table S1. Editing strategies used to generate isogenic models of HD in HEK 293T cells and hiPSCs.

HEK 293T					
	Endonuclease	**sgRNA**	**Donor template**	**HDR efficiency (%)**	**Results**
1	Cas9 nickase, plasmid	Pair: HTT_sg1 and HTT_sg4	ssODN (silent mutation in a PAM)	0	Indel mutations
2	Cas9 nickase, plasmid	HTT_sg4	ssODN (silent mutation in a PAM)	0	Indel mutations
3	Cas9 wt, plasmid	HTT_sg4	ssODN (silent mutation in a PAM)	0	Indel mutations
4	Cas9 wt, plasmid	HTT_sg4	Plasmid with exon 1 of the *HTT* gene	0	Indel mutations
5	Cas9 wt, plasmid	HTT_sg3	Plasmid with exon 1 of the HTT gene	7/109 (6.4%)	Homo- and heterozygous monoclones, all with indel mut in a cut site
6	Cas9 wt, protein	HTT_sg3, RNA	Plasmid with exon 1 of the *HTT* gene (silent mutation in a PAM)	41- 15/70 (21.42%) 53- 14/83 (16.87%) 84- 13/71 (18.31%)	Homozygous monoclones with 41, 53 and 84 CAG; (silent mutation in a PAM) and heterozygous monoclones all with indel mut in a cut site
hiPSCs					
1	Cas9 nickase, plasmid	Pair: HTT_sg1 and HTT_sg4	ssODN; 10 CAG, 180 nt	0	Low electroporation efficiency, apoptosis, CAG excision, strand rejoining
2	Cas9 nickase, plasmid	Pair: HTT_sg1 and HTT_sg4	ssODN (silent mutation in a PAM), longer arms, 300 nt	0	Low electroporation efficiency, apoptosis, CAG

					excision, strand rejoining
3	Cas9 wt, protein	HTT_sg3, RNA	ssODN 10 CAG, 18 0nt	0.61% (10mut/19CAG); 1.23% (19/19CAG); 1.23% (109/109 CAG)	163 monoclones in total
4	Cas9 wt, protein	HTT_sg3, RNA	Plasmid with exon 1 of the *HTT* gene (19 CAG; silent mutation in a PAM)	6% (19/19CAG)	131 monoclones in total

Supplemental Table S2. Summary of capture statistics for whole-exome sequencing.

Sample ID	**Mean Read Length**	**Total Reads (in million)**	**After Removing Identical Reads (in million)**	**Unique (%)**	**Mapped Reads (in million)**	**Mapping (%)**
ND42222	99	105,3	90,674	86,11	89,401	98,6
C37	99	123,391	104,87	84,99	102,94	98,16
C39	99	95,493	83,098	87,02	81,936	98,6
C31.9	99	149,826	122,902	82,03	120,54	98,08

Supplemental Table S3. Sequence Variants in the gene-corrected hiPSC clones by whole-exome sequence analysis

	C37	**C39**	**C31.9**
Total*	86	54	88
3_prime_UTR	0	0	0
3_prime_UTR_intronic	0	0	0
5_prime_UTR	3	1	1
5_prime_UTR_intronic	0	0	0
downstream_gene	0	0	0
essential_splice_site	0	0	0
initiator_codon	0	0	0
intronic	22	22	24
kozak_sequence	0	0	0
missense	31	15	29

non_coding_exonic	0	0	1
non_coding_intronic	0	0	1
splice_region	6	5	5
stop_gained	0	0	0
stop_lost	0	0	0
stop_retained	0	0	0
synonymous	24	11	27
upstream_gene	0	0	0

*OFA>0.85

Supplemental Table S4. Sequence variants present in all edited clones C37, C39, C31.9 detected by whole-exome sequencing.

Chr	Position	Gene	Ref	Alt	Consequence
chr1	220789362	MARK1	T	A	intronic
chr2	131220864	POTEI	T	A	missense
chr2	214012405	IKZF2	A	C	intronic
chr3	195506446	MUC4	G	T	missense
chr3	195510582	MUC4	A	C	synonymous
chr5	80756855	SSBP2	A	T	intronic
chr5	87502325	TMEM161B	G	A	intronic
chr6	18134021	TPMT	C	A	intronic
chr9	34725368	FAM205A	A	G	synonymous
chr10	29580942	LYZL1	C	T	splice_region
chr10	29580944	LYZL1	C	A	intronic
chr10	29580961	LYZL1	C	T	intronic
chr10	46321555	AGAP4	G	A	synonymous
chr12	10571716	KLRC3	A	T	intronic
chr12	10571716	AC068775.1	A	T	intronic
chr12	27826780	PPFIBP1	T	G	intronic
chr15	30906259	GOLGA8H	G	T	splice_region
chr15	32743481	GOLGA8O	C	T	intronic
chr15	69715488	AC027237.1	C	T	intronic
chr15	69715488	KIF23	C	T	intronic
chr16	71805160	AP1G1	G	A	chr16
chr19	7051376	MBD3L2	G	A	missense
chr19	43860251	CD177	G	A	chr19
chr19	43860255	CD177	T	G	missense

chr22	21797094	HIC2	C	G	5_prime_UTR
chr22	24300660	GSTT2B	G	C	intronic

Supplemental Table S5. DNA and RNA oligonucleotides used for the generation of sgRNAs

Oligonucleotide	**Sequence (5′-3′)**	**Description**
HTT_sg1s	**CACC**GCTGCTGCTGCTGCTGGA	oligo for Cas9_HTT.sg1 plasmid construction
HTT_sg1a	**AAAC**TCCAGCAGCAGCAGCAGC	oligo for Cas9_HTT.sg1 plasmid construction
HTT_sg3s	**CACC**GGAAGGACTTGAGGGACTCGA	oligo for Cas9_HTT.sg3 plasmid construction
HTT_sg3a	**AAAC**TCGAGTCCCTCAAGTCCTTCC	oligo for Cas9_HTT.sg3 plasmid construction
HTT_sg4s	**CACC**GGCTTCCTCAGCCGCCGCCGC	oligo for Cas9_HTT.sg4 plasmid construction
HTT_sg4a	**AAAC**GCGGCGGCGGCTGAGGAAGCC	oligo for Cas9_HTT.sg4 plasmid construction
HTT_sg3 RNA	IDT	tracrRNA, RNP strategy
HTT_sg3 crRNA	GAAGGACUUGAGGGACUCGA	crRNA, RNP strategy
ssODN	TGCCGTGCCGGGCGGGAGACCGCCATGGCGACCCT GGAAAAGCTGATGAAGGCCTTCGAGTCCCTCAAGT CCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG CAACAGCCGCCACCGCCGCCGCCGCCGCCGCCGCC TCCTCAGCTTCCTCAGCCGCCGCCGCAGGCACAGCC GCTG	donor template for HDR

Supplemental Table S6. DNA oligonucleotides used as primers for PCR, RT-qPCR and directed mutagenesis.

Name	Forward (5′-3′)	Reverse (5′-3′)	Method
HD1	CCGCTCAGGTTCTGCTTTTA	GGCTGAGGCAGCAGCGGCTG	PCR, seq
−17f and Exon2r*	GAGCCGCTGCACCGAC	CTGACAGACTGTGCCACTATGTTT	PCR
2805f and 2959r*	GATTTTGGCAGTTCTGTTCACG	ATAAACTGAGGCCCATGCATG	PCR
Fsp2 and Rsp2	CTGCACCGACCGTGAGTT	CAAGGGAAGACCCAAGTGAG	PCR
HD 3′CAG	CGACAGCGAGTCAGTGATTG	ACCACTCTGGCTTCACAAGG	RT-qPCR
SOX2	CAAAAATGGCCATGCAGGTT	AGTTGGGATCGAACAAAAGCTATT	RT-qPCR
NANOG	TTTGGAAGCTGCTGGGGAAG	GATGGGAGGAGGGGAGAGGA	RT-qPCR
OCT3/4	AGTTTGTGCCAGGGTTTTTG	ACTTCACCTTCCCTCCAACC	RT-qPCR
Beta actin	TGAGAGGGAAATCGTGCGTG	TGCTTGCTGATCCACATCTGC	RT-qPCR
GAPDH	GAAGGTGAAGGTCGGAGTC	GAAGATGGTGATGGGATTTC	RT-qPCR
mutHDg3	GGAAAAGCTGATGAAGGCGTTCGAGTCCCTCAAGTC	GGACTTGAGGGACTCGAACGCCTTCATCAGCTTTTC	Directed mutagenesis

*Sequences of primers are from Sathasivam K. et al., (2013) Proc Natl Acad Sci U S A, 110, 2366–2370

Supplemental Table S7. Antibodies used for immunocytochemistry (ICC) and western blotting (WB)

ICC	Antibody	Dilution	Company Cat # and RRID
Primary antibody (pluripotency markers)	Rabbit anti-OCT4	1:200	ThermoFisher Cat# PA5-27438 RRID: AB_2544914
	Rabbit anti-NANOG	1:200	Cell Signaling Cat#4903 RRID: AB_10559205
	Mouse anti-TRA-1-60	1:100	Millipore Cat# MAB4360 RRID: AB_2119183
	Mouse anti-TRA-1-80	1:100	ThermoFisher Cat#MA1-024 RRID: AB_2536706
Secondary antibody	Donkey Anti-Rabbit Alexa Fluor 488	1:1000	Jackson ImmunoResearch, West Grove, PA, USA Cat#711-546-152

			RRID: AB_2340619
Secondary antibody	Donkey Anti-Mouse Alexa Fluor 594	1:1000	Jackson ImmunoResearch Cat#715-586-151 RRID: AB_2340858
WB	**Antibody**	**Dilution**	**Company Cat # and RRID**
Primary antibody	Rabbit anti-huntingtin [EPR5526]	1:1000	Abcam, Cambridge, UK Cat#ab109115
Primary antibody	Rabbit anti-plectin	1:1000	Cell Signaling, Leiden, NED Cat#12254 RRID:AB_2797858
Secondary antibody	Anti-rabbit HRP-conjugate	1:2000	Jackson ImmunoResearch Cat# 711-035-152 RRID: AB_10015282