Rozprawa doktorska

Charakterystyka molekularna transkryptów *HTT* i *ATXN3* w kontekście ich roli w patogenezie i użycia jako cele w terapii chorób poliglutaminowych HD i SCA3

mgr Paweł Joachimiak

Praca wykonana w Zakładzie Biotechnologii Medycznej ICHB PAN
pod opieką dr hab. Agnieszki Fiszer, prof. ICHB PAN

Poznań, 2023

Składam serdeczne podziękowania

Mojej Promotor ***dr hab. Agnieszce Fiszer, Prof. ICHB***

za opiekę naukową, cenne uwagi merytoryczne, okazaną życzliwość i wyrozumiałość. Za poświęcony czas oraz świadomość, że mogę przyjść do Niej z każdym swoim problemem czy pytaniem.

Dr Adamowi Ciesiołce

Za przekazaną wiedzę, naukę projektowania eksperymentów oraz za dyskusje naukowe oraz światopoglądowe.

Koleżankom i Kolegom z Zakładu Biotechnologii Medycznej oraz Zakładu Inżynierii Genomowej

za pomoc i cenne rady oraz stworzenie miłej atmosfery w pracy.

Rodzicom i Najbliższym*, a w szczególności* ***Mamie i Babci***

za wychowanie, wyrozumiałość i niekończące się wsparcie

oraz ***Żanecie***

za okazaną wiarę we mnie, wsparcie, pomoc oraz za naukę by cieszyć się nawet z najmniejszych swoich sukcesów.

Spis treści

1. Wykaz publikacji

1.1. Publikacje wchodzące w skład rozprawy doktorskiej

- A. Ciesiołka*, A. Stroynowska-Czerwińska*, **P. Joachimiak**, A. Ciołak, E. Kozłowska, M. Michalak, M. Dąbrowska, M. Olejniczak, K. Raczyńska, D. Zielińska, M. Woźna-Wysocka, W. Krzyżosiak, A. Fiszer. *„Artificial miRNAs targeting CAG repeat expansion in ORFs cause rapid deadenylation and translation inhibition of mutant transcripts”.* Cell Mol Life Sci. 2021 Feb;78(4):1577-1596. doi: 10.1007/s00018-020-03596-7.
*równorzędni współautorzy

IF (2021): **9.234** 5-letni IF: **10.001** Punkty MNiSW: **140**

- **P. Joachimiak**, A. Ciesiołka, G. Figura, A. Fiszer. *„Implications of Poly(A) Tail Processing in Repeat Expansion Diseases”*. Cells. 2022 Feb;11(4):677.doi: 10.3390/cells11040677.

IF (2021): **7.666** 5-letni IF: **7.677** Punkty MNiSW: **140**

- **P. Joachimiak**, A. Ciesiołka, E. Kozłowska, P. M. Świtoński, G. Figura, A. Ciołak, G. Adamek, M. Surdyka, Ż. Kalinowska-Pośka, M. Figiel, N. S. Caron, M. R. Hayden, A. Fiszer. *„Allele-specific quantitation of ATXN3 and HTT transcripts in polyQ disease models”*. BMC Biol. 2023 Feb;21(1):17. doi: 10.1186/s12915-023-01515-3.

IF (2021): **7.364** 5-letni IF: **8.641** Punkty MNiSW: **140**

1.2. Publikacje niewchodzące w skład rozprawy doktorskiej

- B. Tymoniuk*, P. Zmora*, J. Latowska, A. Grabowska, A. Ciesiołka, **P. Joachimiak**, G. Figura, M. Borowiec, K. Rolle, L. Handschuh, A. Fiszer. *„Genetic tests based on the RT-PCR reaction in the diagnostics of SARS-CoV-2 infection”*. Przegl Epidemiol. 2021;75(1):14-26. doi: 10.32394/pe.75.02.
*równorzędni współautorzy

Punkty MNiSW: **40**

2. Streszczenie pracy

Choroby poliglutaminowe (poliQ) są to choroby neurodegeneracyjne, w większości autosomalne dominujące. Są spowodowane mutacją, która polega na wydłużeniu ciągu powtórzeń CAG w otwartych ramkach odczytu (ang. *open reading frame*, ORF) konkretnych genów, co skutkuje obecnością wielu reszt glutaminy w kodowanych białkach. Pacjenci zazwyczaj posiadają dwa allele danego genu – allel normalny (WT) oraz allel zmutowany (MUT). Dla większości chorób poliQ, allel normalny charakteryzuje się liczbą powtórzeń CAG z zakresu 5-30, natomiast liczba powtórzeń CAG w allelu zmutowanym dla większości chorób zaczyna się od ok. 39. Dwie najlepiej poznane i najszerzej opisane choroby poliQ to choroba Huntingtona (ang. *Huntington's disease*, HD) oraz ataksja rdzeniowo-móżdżkowa typu 3 (ang. *spinocerebellar ataxia type 3*, SCA3). Mimo ciągle postępującej wiedzy na temat tych chorób, cały czas nie udało się opracować skutecznej terapii.

Większość opracowań literaturowych dotyczących chorób poliQ porusza role zmutowanych białek, jako głównych czynników patogennych. W przypadku transkryptów poliQ wykazano, że mogą stanowić dobry cel terapeutyczny dla potencjalnych strategii dążących do obniżenia ekspresji zmutowanych genów. Jednakże, rola transkryptów w patogenezie chorób poliQ nie została jeszcze dobrze zbadana. Stąd też głównym celem niniejszej pracy doktorskiej było poznanie cech transkryptów związanych z chorobami poliQ, które mają znaczenie w patogenezie oraz dla projektowania strategii terapeutycznych.

W pierwszej kolejności podjąłem aspekt terapeutyczny, a konkretnie brałem udział w poznaniu mechanizmu prowadzącego do allelo-selektywnego wyciszania ekspresji zmutowanych genów poliQ, na przykładzie oligonukleotydu A2 celującego w ciąg powtórzeń CAG w sekwencji mRNA huntingtyny (*HTT*). Wyniki badań pozwoliły wykazać udział w mechanizmie wyciszania takich procesów jak inhibicja translacji czy deadenylacja transkryptu, a także pozwoliły określić w nim rolę białka AGO2. Następnie, poruszając kontekst zaangażowania zmutowanych transkryptów w patogenezę chorób poliQ, przeanalizowałem dostępne informacje i dane dotyczące poliadenylacji i alternatywnej poliadenylacji (ang. *alternative polyadenylation*, APA) tych mRNA i zaproponowałem dalsze potencjalne perspektywy w tych badaniach. Kolejno,

opracowałem podejście eksperymentalne służące do ilościowej analizy endogennych transkryptów ataksyny-3 (*ATXN3*) oraz *HTT*. Wykorzystuje ono technikę emulsyjnego PCR (ang. *droplet digital PCR,* ddPCR) oraz warianty SNP (ang. *single nucleotide polymorphism*) pozwalające na rozróżnienie transkryptu WT od transkryptu MUT. Następnie wykorzystałem w/w metodę do precyzyjnego określenia stosunku procentowego endogennych transkryptów (WT/MUT) w różnych liniach komórkowych wyprowadzonych od pacjentów HD i SCA3 oraz w tkankach pobranych z modelu mysiego HD. Wykazałem, że wraz z różnicowaniem neuronalnym może zmieniać się stosunek procentowy alleli WT/MUT, a także ulega zmianie łączna liczba (WT+MUT) badanych transkryptów przypadająca na jedną komórkę. Ponadto, opisane podejście eksperymentalne zostało wykorzystane do oceny efektywności allelo-selektywnej strategii terapeutycznej dla chorób poliQ, jak i do weryfikacji czy metody inżynierii genetycznej wpływają na ekspresję modyfikowanych alleli.

Podsumowując, moje badania wniosły wkład do lepszego poznania zmutowanych transkryptów, zarówno w kontekście rozwoju choroby oraz jako celów molekularnych w strategiach terapeutycznych dla chorób poliQ.

3. Abstract

Polyglutamine diseases (polyQ) are a group of neurodegenerative disorders. They are caused by the expansion of a CAG repeat tract in the coding region of specific genes, which results in an expanded tract of glutamines in protein sequence. Due to an autosomal dominant inheritance pattern, most of polyQ disease patients carry two alleles of specific gene – normal allele (WT), and mutant allele (MUT). For most polyQ disorders, normal allele is characterized by the range of 5-30 CAG repeats, while mutant allele is usually possesses a tract of > 39 CAGs. Huntington's disease (HD) and spinocerebellar-ataxia type 3 (SCA3) are two best described polyQ diseases. Although, our knowledge about polyQ diseases is still rising, there is no efficient treatment for polyQ patients.

Generally, a mutant polyQ protein is recognized as the main pathogenic factor in polyQ diseases. In case of polyQ transcripts, it was observed that they are good therapeutic targets for strategies aimed at the downregulation of mutant polyQ genes. On the other hand, the exact role of mutant transcripts in pathogenic mechanisms in polyQ diseases is being unraveled. Hence, the main goal of my PhD thesis was to become acquainted with polyQ transcripts' characteristics which can affect both, the pathogenesis of polyQ diseases, and the design of therapeutic approaches.

At the beginning of my PhD studies I investigated therapeutic aspect related to polyQ disorders. I participated in unraveling the mechanism leading to allele-selective silencing of mutant polyQ genes, using a specific A2 oligonucleotide targeting CAG repeat tract located within huntingtin (*HTT*) transcript sequence. Obtained results allowed for indicating the occurrence of translation inhibition and transcript deadenylation, as well as for determining the role of AGO2, in the investigated mechanism. Next, I started exploring implications of mutant polyQ transcripts in the pathogenesis of polyQ diseases. I reviewed results from studies exploring polyadenylation and alternative polyadenylation (APA) processes in polyQ transcripts, and I proposed future perspectives in this topic. Then, I developed an experimental approach designed for quantitative analysis of endogenous ataxin-3 (*ATXN3*) and *HTT* transcripts. This approach utilizes droplet digital PCR (ddPCR) technology and heterozygous SNP variants

present in sequences of these transcripts, which allow for discrimination between WT and MUT transcripts. I used this approach for precise, accurate and quantitative determination of the WT/MUT allele ratio in a set of HD and SCA3 patient-derived cell lines, as well as in brain tissues derived from HD mouse model. My experiments demonstrated that neural differentiation can impact the WT/MUT allele ratio, as well as the total number of transcripts (WT+MUT) per cell. Mover, this approach was used to assess the efficacy of allele-selective therapeutic strategy, and to verify whether genome engineering techniques affect the expression of targeted alleles. To conclude, my research contributed to a better understanding of mutant polyQ transcripts in the context of both, the pathogenesis of polyQ disorders, and as molecular targets in therapeutic strategies.

4. Wprowadzenie

4.1. Choroby poliglutaminowe (poliQ)

Choroby poliQ należą do grupy neurodegeneracyjnych chorób wywołanych ekspansją powtórzeń tandemowych zlokalizowanych w określonych genach. W chorobach poliQ dochodzi do mutacji dynamicznej ciągu powtórzeń CAG w otwartych ramkach odczytu (ang. *open reading frame*, ORF) konkretnych genów, w związku z czym w sekwencji kodowanego białka mutacja skutkuje długą sekwencją występujących po sobie glutamin. W większości są to choroby dziedziczone autosomalnie dominująco, a pacjenci przeważnie posiadają dwa allele genu: allel normalny (WT) oraz allel zmutowany (MUT). Opisanych zostało 9 chorób poliQ: ataksje rdzeniowo-móżdżkowe typu 1, 2, 3, 6, 7 oraz 17 (ang. *spinocerebellar ataxia*, SCA), choroba Huntingtona (ang. *Huntington's disease*, HD), zanik zębatoczerwienny pallidoniskowzgórzowy (ang. *dentatorubral-pallidoluysian atrophy*, DRPLA) oraz jedyna choroba poliQ niedziedziczona autosomalnie dominująco, związany z chromosomem X – rdzeniowo-opuszkowy zanik mięśni (ang. *spinal and bulbar muscular atrophy*, SBMA). Spośród chorób poliQ, najczęściej występującymi w populacji i najlepiej opisanymi w literaturze są HD oraz SCA3.

Pierwsze objawy chorób poliQ występują z reguły dopiero w trzeciej lub czwartej dekadzie życia pacjentów. W zależności od choroby, różna jest liczba powtórzeń CAG niezbędna do wystąpienia patologii, natomiast wspólną cechą genów zaangażowanych w choroby poliQ jest to, że im większa jest liczba powtórzeń CAG w zmutowanym genie, tym wcześniej ujawnia się dana choroba i silniejsze są jej objawy [1, 2]. Dodatkowo, wydłużenie ciągu CAG jest mutacją somatycznie niestabilną, co może prowadzić do zmiany długości ciągu w trakcie życia pacjenta [3]. Pomimo wspólnej, charakterystycznej cechy genetycznej – mutacji powodującej wydłużenie ciągu powtórzeń – zaangażowane w choroby poliQ geny nie są ze sobą ściśle powiązane, a białka poliQ pełnią różne role w komórkach [4]. Ponadto, ekspresja zmutowanych genów zachodzi w większości komórek i tkanek organizmu ludzkiego, ale efekty patologiczne choroby dotyczą konkretnych typów komórek i tkanek, głównie nerwowych [1]. Głównym czynnikiem patogennym chorób poliQ jest zmutowane białko, które może wywoływać wiele zaburzeń w komórce [1, 5, 6]. Białka poliQ są nieprawidłowo zwijane i mogą tworzyć agregaty oraz sekwestrować inne, kluczowe dla procesów komórkowych białka.

Agregaty są często są zatrzymywane w jądrze komórkowym, co może niekorzystnie wpływać na ekspresję genów czy organizację i funkcje samego jądra komórkowego. Ponadto, zmutowane białka poliQ mogą uczestniczyć w procesach, w których zaangażowane są prawidłowe warianty tych białek, ale pełnić te funkcje w sposób prowadzący do zaburzeń.

Niestety, nie istnieje obecnie żadna skuteczna metoda leczenia chorób poliQ – leczenie jest objawowe, mające na celu jedynie poprawienie jakości życia pacjentów, a zgon następuje najczęściej w ciągu kilkunastu lat po wystąpieniu pierwszych objawów. Spośród strategii terapeutycznych opracowywanych dla chorób poliQ, duże nadzieje są wiązane z tymi mającymi na celu hamowanie ekspresji zmutowanego genu. Wykorzystują one różne podejścia: interferencję RNA (ang. *RNA interference*, RNAi) [7, 8], zastosowanie oligonukleotydów antysensowych (ang. *antisense oligonucleotides*, ASO) [9–11] czy narzędzi terapii genowych jak CRISPR-Cas9 [12].

4.1.1. Choroba Huntingtona (HD)

HD wywoływana jest mutacją w genie kodującym huntingtynę (*HTT*). Jest to duży gen – jego *locus* jest długości 180 tysięcy par zasad (pz) i składa się z 67 egzonów, a ciąg powtórzeń CAG znajduje się na samym jego początku, w egzonie 1, ok. 50 pz od kodonu START. Allel normalny *HTT* posiada liczbę powtórzeń z zakresu 6-35 CAG (najczęściej ok. 17 CAG), natomiast allel zmutowany charakteryzuje się liczbą powtórzeń z przedziału 40-250 powtórzeń CAG (przeważnie od 40 do 50 CAG) [13]. Pierwsze objawy HD występują zazwyczaj ok. 35-40 roku życia, ale istnieje także młodzieńcza postać choroby wywołana dużą liczbą powtórzeń (> 60 CAG), która może objawiać się już w pierwszych dwóch latach życia [14, 15]. Choroba atakuje zwłaszcza neurony prążkowia i kory mózgu, a głównymi objawami są mimowolne ruchy nóg, ramion, twarzy (stąd też inna nazwa choroby – pląsawica Huntingtona), oraz zaburzenia psychiczne i emocjonalne [1, 13, 15]. Allel WT koduje białko huntingtynę o masie ok. 350 kDa, którego rola nie jest jeszcze w pełni poznana. Wiadomo, że huntingtyna pełni istotne funkcje na etapie neurogenezy w rozwoju embrionalnym – u myszy delecja jej genu na tym etapie jest letalna [16]. Ponadto, jej ważną rolą jest transport istotnego dla neuronów prążkowia i kory mózgowej czynnika pochodzenia mózgowego (ang. *brain-derived neurotrophic factor*, BDNF) [17]. Dodatkowo, huntingtyna odpowiada za transport organelli i pęcherzyków

wzdłuż aksonów oraz za wiązanie czynników transkrypcyjnych [18]. Mutacja prowadzi natomiast m.in. do deregulacji transkrypcji, zaburzeń transportu pęcherzykowego oraz dysfunkcji mitochondriów [15]. W ostatnich latach coraz większą uwagę naukowców zajmujących się HD przykuwa skrócona forma białka – HTTegzon1 – powstająca w wyniku zaburzeń splicingu zmutowanego transkryptu [19]. Forma ta jest wysoce patogenna i mająca szczególną skłonność do agregacji [20, 21].

4.1.2. Ataksja rdzeniowo-móżdżkowa typu 3 (SCA3)

SCA3, znana również pod nazwą choroby Machado-Josepha, powodowana jest mutacją w genie *ATXN3,* kodującym ataksynę-3, który składa się z 11 egzonów, a ciąg powtórzeń zlokalizowany jest bliżej końca 3' genu – w egzonie 10. Allel normalny *ATXN3* charakteryzuje się ciągiem powtórzeń w zakresie 12-44 CAG (najczęściej ok. 18 CAG), natomiast allel zmutowany posiada liczbę powtórzeń z zakresu 60-87 CAG [13]. W przypadku tej choroby degeneracja neuronów zachodzi głównie w móżdżku w jądrze zębatym, a także m.in. w pniu mózgu, nerwach czaszkowych oraz w rdzeniu kręgowym [22]. Wśród objawów choroby wymienia się ataksję, zaburzenia koordynacji ruchowej, zaburzenia mowy, dystonię, parkinsonizm, a także zaburzenia psychiczne, zaburzenia snu oraz depresję [22]. Ataksyna-3 jest białkiem o masie 42 kDa zawierającym N-terminalną domenę *Josephin* o aktywności deubikwitynazy oraz C-terminalny rejon posiadający trzy motywy oddziałujące z ubikwityną (ang. *ubiquitin-interacting motifs*, UIM). Taka budowa ataksyny-3 determinuje jej rolę w systemie ubikwityna-proteasom (ang. *ubiquitin-proteasome system*, UPS) [23]. Ataksyna-3 bierze także udział w procesie autofagii, czyli w procesie degradacji dużych białek (większych niż w przypadku degradacji w proteasomie), kompleksów białkowych lub organelli [24]. A zatem, ataksyna-3 jest ważnym elementem swoistej komórkowej kontroli jakości produkowanych białek [25]. Dodatkowo, ataksyna-3 pełni role w regulacji transkrypcji [26] oraz w mechanizmach naprawy DNA [22, 27]. Obecność zmutowanego białka prowadzi m.in. do dysfunkcji UPS, zaburzeń autofagii, dysfunkcji mitochondriów oraz deregulacji transkrypcji [28, 29].

4.2. Rola transkryptów w patogenezie chorób poliQ

Od wielu lat za główny czynnik patogenny chorób poliQ uznaje się zmutowane białka [30–32], natomiast rola transkryptów poliQ w patogenezie długo była nieznana lub pomijana. Pierwsze wzmianki sugerujące wpływ toksyczności RNA na patogenezę chorób poliQ odnosiły się do podobieństwa pomiędzy powtórzeniami CAG a CUG, które już wcześniej opisano (w badaniach przeprowadzonych w ICHB PAN), jako mające tendencję do formowania drugorzędowych struktur przypominających spinkę do włosów (ang. *hairpin*) [33–35]. Natomiast pierwsze istotne dowody zostały dostarczone dzięki eksperymentom wykonanym na modelu *Drosophila melanogaster*. Wykazano w nich, że obecność transkryptów z wydłużonymi ciągami powtórzeń CAG, w wariancie nie ulegającym translacji, prowadzi do neurodegeneracji [36]. Od tamtego czasu, liczba dowodów wskazujących na przynajmniej częściową rolę zmutowanych transkryptów w patogenezie chorób poliQ stale rośnie. Wśród procesów związanych ze zmutowanymi RNA można wyróżnić: sekwestrację białek wiążących RNA (ang. *RNA-binding proteins*, RBPs), tworzenie skupień (*foci*), aktywację stresu jąderkowego, powstawanie małych RNA CAG (ang. *small CAG RNAs*, sCAG), RAN translację (ang. *repeat-associated non-AUG translation*), zaburzenia splicingu oraz alternatywną poliadenylację (ang. *alternative polyadenylation*, APA).

Przykładowym białkiem wiązanym przez zmutowane ciągi powtórzeń CAG jest nukleolina (ang. *nucleolin*, NCL) – białko występujące w jąderku i biorące udział w procesach komórkowych takich jak: transkrypcja, stabilizacja mRNA, biogeneza rybosomów oraz translacja [37–40]. Deregulacja NCL została potwierdzona jako jeden z mechanizmów związanych z patogenezą w nowotworach [41, 42] i chorobach neurodegeneracyjnych [43, 44]. Transkrypty z wydłużonym ciągiem powtórzeń CAG sekwestrują NCL, co prowadzi do inhibicji transkrypcji rRNA, a finalnie powoduje stres jąderkowy oraz apoptozę zależną od p53 [45, 46]. W niedawnym badaniu [47] potwierdzono, że aminokwasowe reszty NCL odpowiedzialne za wiązanie ciągów CAG nie biorą udziału w naturalnych funkcjach badanego białka. Tym samym, zaproponowano zmutowaną wersję NCL, która nie wiąże powtórzeń CAG, jako potencjalny czynnik terapeutyczny do zmniejszenia stresu jąderkowego wywołanego wydłużonymi ciągami CAG w terapiach chorób poliQ.

Następnym przykładem toksyczności transkryptów poliQ jest powstawanie sCAG. Rybonukleaza Dicer rozpoznaje i tnie transkrypty z długimi ciągami CAG tworząc sCAG, ok. 20 nt RNA [48] – jest to mechanizm odkryty w ICHB PAN. Mogą one obniżać ekspresję genów, do których mają w pełni lub częściowo komplementarną sekwencję, na drodze mechanizmu RNAi. Takie działanie może prowadzić do niezamierzonych zmian w funkcjonowaniu procesów komórkowych. sCAG mają szczególne powinowactwo do komplementarnych powtórzeń CUG. Przykładem może być obniżenie ekspresji *NUDT16*. Białko NUDT16 należy do rodziny hydrolaz Nudix (ang. *nucleoside dephosphatase linked moiety X*), zlokalizowane jest w jądrze komórkowym i posiada aktywność difosfatazy deoksyinozyny [49]. Jego rolą jest usuwanie niechcianych nukleotydów z puli komórkowej, np. deaminowanej puryny, której wbudowanie do nici DNA prowadzi do mutacji [50]. Poziom ekspresji *NUDT16* jest obniżony w modelach chorób poliQ - transkrypt *NUDT16* posiada w swojej sekwencji powtórzenia CUG, z którymi sCAG tworzą heterodupleksy RNA, co z kolei prowadzi do wyciszenia ekspresji genu w procesie zależnym od kompleksu RISC (ang. *RNA-induced silencing complex*) [51]. Skutkuje to powstawaniem uszkodzeń DNA oraz apoptozą komórek. Obecność sCAG potwierdzono także w tkankach mózgowych pobranych od pacjentów HD [52].

W kontekście transkryptów poliQ, zaburzenia splicingu oraz wpływ APA najlepiej opisano na przykładzie *HTT*. Jeszcze w poprzednim wieku wygenerowano modele z wysoce toksyczną, skróconą formą białka mającą szczególną tendencję do agregacji – HTTegzon1 [20, 21]. W późniejszych badaniach wykazano, że forma ta może powstawać endogennie w wyniku translacji transkryptu *HTT1a*, który z kolei powstaje w skutek zaburzeń splicingu zmutowanego transkryptu [19, 53]. Transkrypt ten posiada w swojej sekwencji cały egzon 1 *HTT* wraz z wydłużonym ciągiem powtórzeń CAG oraz część intronu 1. Aby mógł powstać dojrzały transkrypt *HTT1a*, jedno z dwóch kryptycznych miejsc poliadenylacji zlokalizowanych w intronie 1 musi zostać aktywowane – w przypadku ludzkiej *HTT* miejsca te są zlokalizowane 2710 oraz 7327 pz w sekwencji intronu 1, a w przypadku mysiej *HTT* jest to 680 oraz 1145 pz od końca 5’ intronu 1 [19]. Ilość powstającego transkryptu *HTT1a* jest ściśle skorelowana z długością powtórzeń CAG – dłuższy ciąg powoduje powstawanie jego większej ilości [19]. Początkowo sądzono, że za zaburzeniem splicingu stoi wiązanie czynnika transkrypcyjnego SRSF6

przez położony w pobliżu wydłużony ciąg powtórzeń [19], jednakże późniejsze doświadczenia na myszach wykazały, że SRSF6 nie wpływa na powstawanie *HTT1a* [54]. Niedawne badania przeprowadzone na myszach HD YAC128 (są to myszy transgeniczne z pełnej długości sekwencją ludzkiej *HTT*) wskazują, że forma *HTT1a* powstaje we wszystkich badanych rejonach mózgu, przy czym rejonem wykazującym się najwyższym poziomem tego transkryptu jest móżdżek [55]. Transkrypt *HTT1a* może gromadzić się w jądrze gdzie kolokalizuje z pełnej długości mRNA *HTT*, lub może być eksportowany do cytoplazmy, gdzie zachodzi jego translacja do skróconego białka HTTegzon1 [55]. Poza patogenną formą *HTT1a* ściśle związaną z mutacją, opisano także trzy pełnej długości izoformy mRNA normalnej *HTT* różniące się jedynie długością 3'UTR, powstałe w wyniku działania APA [56, 57]. Izoforma długa, mająca 13.7 tysięcy pz i dominująca w komórkach niedzielących się oraz w tkankach jak np. mózg i piersi, wraz z formą krótką, mającą 10.3 kb długości i przeważającą w komórkach aktywnie dzielących się, np. w komórkach mięśniowych i HEK293, uznawane są za główne izoformy. Istnieje także forma pośrednia (12.5 kb długości), która obecna jest zarówno u ludzi jak i u myszy, w materiale kontrolnym jak i w materiale HD [57]. Mimo, że nie powiązano konkretnej izoformy 3'UTR z wystąpieniem objawów HD, to wszystkie wymienione izoformy charakteryzują się różną lokalizacją, długością ogonów poli(A), oraz miejscami wiązania zarówno miRNA, jak i RBP. Cechy te natomiast mogą wpływać na stabilność transkryptów *HTT*, a finalnie na poziom ekspresji *HTT* w konkretnych tkankach i warunkach. Jest to niewątpliwie ciekawy aspekt do dalszych badań w HD, ale także w innych chorobach poliQ, w których podobne efekty i procesy również mogą mieć miejsce.

4.3. Strategie terapeutyczne celowane w transkrypty poliQ

Pomimo stale rosnącej wiedzy na temat charakterystyki chorób poliQ, coraz bardziej zaawansowanej medycyny i kolejnych badań klinicznych, ciągle nie opracowano skutecznej terapii leczenia chorób poliQ, a leczenie obecnie polega jedynie na łagodzeniu objawów. Najwięcej strategii terapeutycznych opracowano dla HD i to na nich głównie skupię się w tym rozdziale. Na przestrzeni lat próbowano różnorodnych podejść terapeutycznych – od zastosowania związków niskocząsteczkowych [58, 59], przez próby zmodyfikowania szlaków komórkowych do zmniejszenia toksyczności

i tworzenia agregatów przez zmutowane białko [60], a kończąc na najbardziej obiecujących strategiach celujących w kwasy nukleinowe (DNA i RNA). Do tych ostatnich zalicza się wykorzystanie narzędzi inżynierii genetycznej jak systemów CRISPR-Cas oraz białek palców cynkowych (ang. *zinc fingers nucleases*, ZFN), a także zastosowanie ASO oraz technologię RNAi.

4.3.1. Terapie allelo-selektywne vs. nieallelo-selektywne

Jednym z pierwszych pytań przy planowaniu strategii terapeutycznej celującej w kwasy nukleinowe jest to, czy wymagane jest aby dana strategia działała allelo-selektywnie. O allelo-selektywności mówimy kiedy terapeutyk wycisza silnie preferencyjnie ekspresję zmutowanego genu pozostawiając ekspresję normalnego allelu niezmienioną lub zmienioną nieznacznie. Takie podejście w przypadku terapii chorób poliQ może być szczególnie pożądane, gdyż istotne wydaje się zachowanie funkcji normalnego białka. W przypadku *HTT* wiemy, że delecja tego genu jest letalna na etapie embriogenezy u myszy [16] oraz, że może negatywnie wpływać na transport aksonalny [61]. I chociaż inne badania, przeprowadzone na makakach, wykazały że częściowe obniżenie ekspresji normalnej *HTT* w dorosłym mózgu - do poziomu ok. 50% - powinno być bezpieczne [62], to nie znamy potencjalnych, długoterminowych skutków obniżenia ekspresji obu alleli *HTT* w mózgach pacjentów HD [63]. Z drugiej jednak strony, terapie allelo-selektywne również mają swoje ograniczenia. Allel normalny od zmutowanego najłatwiej rozróżnić ze względu na długość ciągu CAG, jednakże celowanie w wydłużony ciąg CAG, czyli bezpośrednio w miejsce mutacji, wiąże ze sobą ryzyko, że efektem ubocznym terapii będzie zaburzenie ekspresji innych genów zawierających ciągi powtórzeń CAG [64]. Inną możliwością jest celowanie terapeutycznych oligonukleotydów w rejon zawierający polimorfizm pojedynczego nukleotydu (ang. *single nucleotide polymorphism*, SNP), który byłby obecny jedynie w allelu zmutowanym. Tu jednak ograniczeniem jest fakt, że nie wszyscy pacjenci posiadają dany SNP pozwalający na odróżnienie obu alleli. Przykładowo, wykazano że zaprojektowanie 5 cząsteczek siRNA (ang. *short interfering RNA*) celujących w 3 SNP powinno pozwolić na zaproponowanie skutecznej terapii 75% pacjentów HD [9]. Biorąc pod uwagę zalety i ograniczenia różnych podejść terapeutycznych, w ostatnich latach prowadzono badania zarówno nad podejściem allelo-selektywnym, jak i nieallelo-selektywnym.

4.3.2. Terapie z wykorzystaniem ASO

ASO to krótkie oligonukleotydy o długości ok. 20 nt, które wpływają na ekspresję genu poprzez wiązanie się do mRNA. Wówczas, dochodzi do przecięcia mRNA zależnie od RNazy H, zahamowania translacji lub zmiany wzoru splicingu [65]. Aktywowany mechanizm jest zależny od miejsca wiązania ASO w transkrypcie, jak i od zastosowanych modyfikacji chemicznych oligonukleotydu.

W przypadku terapii HD na najbardziej zaawansowanym etapie badań klinicznych są strategie oparte właśnie o wykorzystanie ASO – badany przez firmę Roche tominersen oraz dwie allelo-selektywne cząsteczki badane przez firmę Wave Life Sciences (WVE-120101 oraz WVE-120102 celujące w dwa różne SNP). Niestety, w 2021 roku obie firmy ogłosiły zakończenie badań na, odpowiednio, III fazie oraz I/II fazie testów klinicznych [66]. W przypadku tominersenu okazało się, że nie wykazywał on skutecznego działania w porównaniu do placebo, a ponadto pacjenci, którym podawany był tominersen, doznawali stosunkowo częstych skutków ubocznych [63]. Z kolei w badaniach Wave Life Sciences nie wykazano statystycznie istotnych różnic w wyciszeniu zmutowanej *HTT* w porównaniu do allelu normalnego [66]. Bazując na niepowodzeniach, obie firmy rozpoczęły nowe badania z wykorzystaniem ASO – firma Roche zmodyfikowała założenia testów klinicznych, natomiast Wave Life Sciences testuje nową cząsteczkę celującą w inny SNP [63]. Trzeba jednak pamiętać, że wciąż jest sporo do poprawy m.in. na polach bezpieczeństwa i sposobu podania ASO. Opierając się na wynikach uzyskanych dla ASO w przypadku innych chorób, można wyciągnąć jednak kilka wniosków. Co najważniejsze, sukces np. nusinersenu (lek na rdzeniowy zanik mięśni, SMA) pokazuje, że ASO z powodzeniem mogą być stosowane w terapiach chorób neurologicznych. Jednakże w przypadku HD testowane ASO były podawane w kilkukrotnie wyższym stężeniu niż wspomniany nusinersen. Ponadto, działanie nusinersenu polega na przywróceniu przynajmniej częściowej ekspresji białka WT, podczas gdy w przypadku HD niezbędne jest znaczące obniżenie ekspresji białka zmutowanego, co jest dużo trudniejsze do osiągnięcia. We wspomnianym wcześniej badaniu [55] zaproponowano system do efektywnego screeningu ASO celujących w ludzkie transkrypty *HTT* z wykorzystaniem mysich embrionalnych fibroblastów (ang. *mouse embryonic fibroblasts*, MEF) wygenerowanych z myszy YAC128. Opisany system może okazać się przydatny

w rozwoju nowych potencjalnych terapii HD. Podsumowując, testowanie terapii z zastosowaniem ASO jest obecnie bardzo zaawansowane i przy odpowiednim dopracowaniu podejście to ma duży potencjał w leczeniu chorób poliQ. Świadczą o tym chociażby dobrze rokujące badania przedkliniczne z wykorzystaniem ASO w terapiach innych chorób poliQ takich jak: SCA1 [67], SCA2 [10], SCA3 [11, 68, 69] oraz SCA7 [70].

4.3.3. Terapie wykorzystujące proces RNAi oraz mechanizm ich działania

Kolejnym przykładem strategii celowania w transkrypty poliQ jest wykorzystanie mechanizmu RNAi. Technologia ta wykorzystuje różniące się budową, czasem działania i sposobem dostarczania do komórek cząsteczki niekodujących RNA (ang. *non-coding RNA*, ncRNA), takie jak chemicznie zsyntetyzowane siRNA czy dostarczane w wektorze shRNA (ang. *short hairpin RNA*), z którego maszyneria komórkowa generuje siRNA. W przypadku podejścia wektorowego, cząsteczki są wprowadzane do docelowych komórek poprzez kodowanie ich sekwencji w wektorze wirusowym - najczęściej jest to AAV (ang. *adeno-associated virus*) – dzięki czemu ekspresja takiej cząsteczki jest stała i długotrwała oraz omijany jest problem „rozcieńczenia" cząsteczek wyciszających wynikający z podziałów komórkowych. Innym przykładem cząsteczek ncRNA wykorzystujących działanie mechanizmu RNAi są celujące w rejon powtórzeń CAG, opracowane m.in. w ICHB PAN i badane w naszym Zakładzie, art-miRNA (ang. *CAG repeat-targeting artificial miRNAs*) [71], których użyteczność była testowana w różnych modelach chorób poliQ. Zawierają one w swojej sekwencji powtórzenia CUG z wprowadzonymi w konkretnych miejscach niesparowaniami do sekwencji docelowej, które upodabniają działanie tych cząsteczek do miRNA (ang. *miRNA-like siRNA*).

Mechanizm działania siRNA polega na tworzeniu w pełni komplementarnego dupleksu pomiędzy siRNA a docelową sekwencją mRNA, a następnie cięciu transkryptu przez AGO2 znajdujące się w kompleksie RISC, w wyniku czego mRNA ulega degradacji [72]. Mechanizm działania miRNA jest bardziej złożony. miRNA prowadzą do wyciszenia ekspresji genu przy udziale takich procesów jak: represja translacji, deadenylacja transkryptu, usuwanie struktury 5' kap (ang. *decapping*) oraz degradacja transkryptu w wyniku działania egzonukleaz [73, 74]. Hamowanie translacji jest złożonym procesem. Pierwotnie sugerowano, że po związaniu miRNA dochodzi do zahamowania formowania kompleksu 80S [75], lub do blokowania fazy elongacji translacji [76, 77]. Jednakże

najnowsze dane sugerują, że represja translacji w wyniku działania miRNA zachodzi na drodze oddziaływań między różnymi białkami prowadzącymi finalnie do blokady skanowania 5'UTR transkryptu przez małą podjednostkę rybosomu (ang. *ribosome scanning*) lub do hamowania inicjacji translacji przez kompetycję z czynnikami inicjacji translacji [78]. Deadenylacja transkryptów przeprowadzana jest dwuetapowo przez kompleksy PAN2-PAN3 oraz CCR4-NOT [79, 80]. Pozbawiony ogona poli(A) transkrypt jest następnie degradowany w kierunku 3'->5' m.in. przez egzosom [81]. Drugi koniec transkryptu jest z kolei poddawany procesowi usuwania 5' kapu, w wyniku którego białko DCP1 rekrutuje egzonukleazę 5'->3' XRN1, która z kolei prowadzi degradację pozbawionego kapu transkryptu [82]. Jednakże zaobserwowano odstępstwa od ogólnie przyjętego modelu oraz znaczące różnice w zakresie aktywowania poszczególnych procesów, tj. represji translacji w porównaniu do degradacji mRNA. Wysokoprzepustowe analizy wykazały, że znacząca większość mRNA regulowanych przez miRNA ulega rozpadowi (ang. *mRNA decay*) [83, 84], natomiast inne badania pokazały, że inhibicja translacji poprzedza deadenylację mRNA [85, 86], a sam rozpad mRNA może mieć miejsce jedynie po represji translacji [87, 88]. Metaanaliza wpływu poszczególnych, wymienionych procesów na wyciszanie mRNA za pomocą miRNA, wykazała że 48% genów było głównie regulowanych przez hamowanie translacji, 29% genów regulowanych było głównie przez rozpad transkryptu, a 23% genów przez oba te procesy [89]. To, który z wymienionych procesów będzie dominował podczas wyciszania za pomocą miRNA, zależy również od lokalizacji miejsca wiązania miRNA w transkrypcie. Większość miejsc wiązania miRNA znajduje się w 3'UTR transkryptów, ale miejsca te są także obecne w ORF [90, 91]. Co ciekawe, w proces wyciszania ekspresji genu za pomocą miRNA celujących w ORF częściej zaangażowana jest inhibicja translacji, a w przypadku miRNA celujących w 3'UTR, większy jest udział procesu rozpadu mRNA [92]. Dokładny mechanizm represji translacji w przypadku celowania miRNA w ORF nie jest poznany, ale sugeruje się zależne od AGO2 przejściowe zatrzymywanie rybosomów na transkrypcie (ang. *ribosome stalling*) [91], lub formowanie alternatywnych kompleksów miRISC zawierających AGO3, zamiast AGO2 [93]. Dlatego też, aby móc opracować jak najbardziej efektywne strategie terapeutyczne w leczeniu m.in. chorób poliQ, istotne jest poznanie dokładnych mechanizmów wyciszania za pomocą siRNA podobnych do miRNA.

5. Cel pracy doktorskiej

Niewiele wiadomo na temat potencjalnej roli transkryptów kodowanych przez zmutowane geny w rozwoju chorób poliQ. Czy są one jedynie przekaźnikiem wadliwej informacji genetycznej z DNA na białko? A może mają także bardziej bezpośrednią funkcję w patogenezie chorób poliQ? Z drugiej strony, w ostatnich latach wykazano, że zmutowane transkrypty poliQ są obiecującymi celami terapeutycznymi w strategiach dążących do obniżenia ekspresji zmutowanych genów poliQ. Natomiast precyzyjne mechanizmy molekularne, potencjalnie zaangażowane w takie strategie, nie są jeszcze dokładnie opisane.

Niniejszym, celem mojej pracy doktorskiej było poznanie cech transkryptów związanych z chorobami poliQ, które mają znaczenie w patogenezie tych chorób oraz dla projektowania strategii terapeutycznych. Kluczowym aspektem moich badań było zastosowanie takich podejść eksperymentalnych, aby móc analizować oba allele badanych transkryptów niezależnie od siebie.

W swojej pracy poruszyłem kilka istotnych wątków wykorzystując różne modele chorób poliQ. Cel badań został osiągnięty poprzez realizację poniższych zadań:

1. Poznanie mechanizmu prowadzącego do allelo-selektywnego wyciszenia ekspresji zmutowanych genów w chorobach poliQ w strategii celowania oligonukleotydów w wydłużony ciąg powtórzeń CAG.
2. Analiza najnowszych informacji dotyczących roli ogonów poli(A) oraz wyboru miejsc poliadenylacji w kontekście chorób poliQ.
3. Opracowanie podejścia eksperymentalnego służącego do ilościowej, allelo-specyficznej analizy endogennych transkryptów związanych z chorobami poliQ.
4. Wykorzystanie w/w podejścia do precyzyjnego określenia stosunku procentowego alleli endogennych transkryptów w różnych liniach komórkowych pacjentów i w modelu mysim.

6. Krótkie omówienie publikacji wchodzących w skład rozprawy doktorskiej

6.1. Artificial miRNAs targeting CAG repeat expansion in ORFs cause rapid deadenylation and translation inhibition of mutant transcripts

A. Ciesiołka, A. Stroynowska-Czerwińska*, **P. Joachimiak**, A. Ciołak, E. Kozłowska, M. Michalak, M. Dąbrowska, M. Olejniczak, K. Raczyńska, D. Zielińska, M. Woźna-Wysocka, W. Krzyżosiak, A. Fiszer. Cell Mol Life Sci. 2021 Feb;78(4):1577-1596. doi: 10.1007/s00018-020-03596-7.*

Aspekt terapeutyczny w kontekście chorób poliQ był pierwszym zagadnieniem, nad którym pracowałem w trakcie swojego doktoratu. W ICHB PAN od lat pracowano nad specyficznymi oligonukleotydami celującymi w ciąg powtórzeń CAG nazywanymi w omawianej publikacji art-miRNA [71, 94–96], z których najbardziej efektywnym okazał się działający allelo-selektywnie oligonukleotyd A2. Charakterystyczną ich cechą jest to, że po związaniu się do wydłużonych ciągów CAG tworzą się niesparowania. Takie wiązanie oligonukleotydów nie prowadzi wówczas do cięcia transkryptu, jak w przypadku siRNA, a działają one w sposób podobny do działania miRNA [94]. Ponadto, zaobserwowano że mechanizm działania wspomnianych art-miRNA polega także na zaangażowaniu białek AGO2 i GW182 oraz na kooperatywnym działaniu kompleksów miRISC, które w liczbie kilku rozpoznają wydłużone ciągi powtórzeń CAG, przez co wyciszają one takie transkrypty bardziej efektywnie niż te z ciągami CAG normalnej długości [97]. Jednakże, nie zostały poznane procesy, które są aktywowane przez tego typu oligonukleotydy. Dlatego też, celem omawianej pracy było dokładnie opisać mechanizm działania art-miRNA. Ważnym aspektem badań tych cząsteczek o potencjale terapeutycznym jest to, że mają one swoje miejsca wiązania w ORF. Inną ciekawą kwestią podjętą w opisywanych badaniach jest różna efektywność działania art-miRNA na poszczególne transkrypty poliQ. Wykazano, że pomimo celowania w ten sam region mutacji (wydłużony ciąg CAG), art-miRNA wykazywały zróżnicowaną aktywność wyciszania dla różnych genów. Dlatego też chcieliśmy określić czynniki, które decydują o tym dlaczego art-miRNA działają z różną efektywnością w zależności od celowanego transkryptu poliQ.

Mój wkład w opisywaną publikację dotyczył poznania mechanizmu wyciszenia *HTT* w odpowiedzi na zastosowanie oligonukleotydu A2. Tę część badań wykonywałem pod

bezpośrednią opieką i ze współudziałem dr Adama Ciesiołki. Pierwszym krokiem było wygenerowanie dwóch linii komórkowych HEK Flp-In T-REx-293 z indukowalną doksycykliną ekspresją egzonu 1 *HTT* zawierającego ciąg 16 lub 98 powtórzeń CAG (nazywane odtąd „16 CAG" oraz „98 CAG"), które odpowiadają dzikiej i zmutowanej *HTT*, odpowiednio. Centralnym elementem obu konstruktów genetycznych służących do wyprowadzenia w/w stabilnych linii komórkowych był promotor dwukierunkowy BI-16. Z jednej strony promotora ekspresji ulegała sekwencja *Firefly* lucyferazy (FLuc), natomiast z drugiej strony promotora ekspresji ulegał wspomniany egzon 1 *HTT* (z 16 CAG lub 98 CAG) oraz sekwencja *Nano* lucyferazy z domeną PEST (NLucP) (Schemat 4A z publikacji). W ten sposób skorzystałem z systemu reporterowego *Dual-Luciferase*®(Promega), który pozwala na obserwację zmian poziomu ekspresji fuzyjnego białka HTT-NLucP w odpowiedzi na oligonukleotydy wyciszające, w odniesieniu do niezmiennej ekspresji białka FLuc użytej do normalizacji. Tak wygenerowane stabilne linie komórkowe posłużyły mi do wszystkich kolejnych eksperymentów.

Pierwszą analizą było zbadanie kinetyki deregulacji transkryptu i białka HTT w odpowiedzi na transfekcje oligonukleotydami wyciszającymi. Linię komórkową 16 CAG oraz linię 98 CAG transfekowałem allelo-selektywnie działającym nukleotydem A2, nieallelo-selektywnie działającym siRNA celującym w sekwencję *HTT* (siHTT) oraz kontrolnym siRNA (siRL). Po 12h od transfekcji, zaindukowałem ekspresję genów przez dodanie doksycykliny i zbierałem materiał komórkowy w określonych punktach czasowych (Rycina 4B w publikacji). Wyizolowane z materiału komórkowego RNA oraz białko zostały użyte, odpowiednio, w reakcji RT-qPCR oraz do pomiaru chemiluminescencji. W wyniku działania siHTT zaobserwowałem znaczący spadek poziomu zarówno transkryptu jak i białka, niezależnie od badanej linii komórkowej (Rycina 4D w publikacji). Wynik ten wskazuje na rolę AGO2 w cięciu obu transkryptów (z normalnej długości oraz z wydłużonym ciągiem CAG) prowadzącym do spadku poziomu obu białek. Natomiast dla transfekcji oligonukleotydem A2, kinetyka deregulacji transkryptu i białka w linii 16 CAG różniła się od tej zaobserwowanej w linii 98 CAG (Rycina 4C w publikacji). Poziom ekspresji zmutowanego transkryptu *HTT* w linii 98 CAG spadł o ok. 50%, podczas gdy poziom transkryptu normalnego w linii 16 CAG obniżył się jedynie nieznacznie w wybranych punktach czasowych. Spadek poziomu

białka w obu liniach był bardziej znaczący, a w dodatku poziom zmutowanego białka spadł jeszcze wyraźniej – o ok. 70%, podczas gdy poziom białka normalnego spadł o ok. 40%. Co istotne, we wczesnych punktach czasowych (do ok. 2h) A2 znacząco obniżył poziom zmutowanego białka, podczas gdy poziom zmutowanego transkryptu pozostawał bez zmian. Sugeruje to, że w opisanym przypadku allelo-selektywnej inhibicji zmutowanej *HTT* hamowanie translacji poprzedza proces rozpadu mRNA.

Jako, że deadenylacja transkryptu jest jednym z procesów zachodzących w wyniku wyciszania ekspresji genów w odpowiedzi na działanie miRNA, w kolejnym eksperymencie chciałem przyjrzeć się bliżej właśnie temu procesowi. Ponownie użyłem linii 16 CAG oraz 98 CAG, które transfekowałem oligonukleotydami A2 oraz kontrolnym siRNA - siRL. By efektywnie móc badać proces deadenylacji, godzinę po indukcji doksycykliną dodałem do komórek aktynomycynę-D w celu zahamowania transkrypcji, a następnie zebrałem materiał komórkowy w trzech punktach czasowych (Schemat 4B w publikacji). Deadenylację, czyli skracanie ogona poli(A) badałem z wykorzystaniem metody *G/I tailing*. W linii komórkowej 98 CAG A2 spowodował znaczącą deadenylację transkryptu *HTT-NLucP* już w ciągu 60 minut od indukcji ekspresji egzogenu (Rycina 4F w publikacji, górny panel). W przypadku transfekcji siRL, podobnego stopnia deadenylacji nie zaobserwowałem nawet po 240 minutach od indukcji, czyli w punkcie czasowym, w którym obserwuje się naturalną deadenylację wynikającą z zatrzymania transkrypcji. Natomiast w przypadku linii 16 CAG nie zaobserwowałem znaczących różnic w tempie deadenylacji pomiędzy komórkami transfekowanymi A2, a tymi transfekowanymi oligonukleotydem kontrolnym (Rycina 4F w publikacji, dolny panel). Podsumowując, A2 spowodowało nagłą deadenylację jedynie zmutowanego transkryptu zawierającego 98 CAG, co sugeruje istotną rolę deadenylacji w represji genów spowodowanej działaniem art-miRNA.

Następnym krokiem było poznanie roli AGO2 w mechanizmie działania art-miRNA poprzez określenie czy aktywność katalityczna AGO2, lub sama jego obecność, jest niezbędna do wyciszenia ekspresji genów za pomocą art-miRNA. W tym celu ponownie użyłem linii komórkowej 98 CAG, która została dodatkowo zmodyfikowana za pomocą technologii CRISPR-Cas9. W wyniku tego powstała linia z delecją AGO2 (AGO2del) oraz z mutacją D597A w sekwencji AGO2 (AGO2mut), która pozbawia białko AGO2

właściwości cięcia transkryptu, natomiast nie zaburza właściwości wiązania krótkich RNA oraz represji translacji [98, 99]. Aby poznać rolę AGO2 w badanym procesie, przeanalizowałem poziomy transkryptu i białka HTT-NLucP w obu w/w liniach komórkowych. Dodatkowo, przeprowadziłem eksperyment typu *rescue* poprzez transfekcję obu linii komórkowych plazmidem posiadającym dziką sekwencję AGO2, aby sprawdzić czy jesteśmy w stanie „cofnąć" efekt wprowadzonych mutacji (Rycina 5B w publikacji). Kiedy obie zmodyfikowane linie były transfekowane siHTT, to nie zaobserwowałem żadnych zmian w poziomach transkryptu oraz białka (Rycina 5C w publikacji). Dopiero ko-transfekcja plazmidem zawierającym dziką sekwencję AGO2 przywróciła pełną funkcjonalność białka AGO2 i widoczny był spadek poziomu transkryptu i białka HTT-NLucP. Natomiast w liniach AGO2del oraz AGO2mut transfekowanych A2 uzyskałem zbliżone obniżenie poziomów transkryptu oraz białka HTT-NLucP, jak w przypadku linii 98 CAG. Wynik ten wskazuje, że art-miRNA A2 obniża ekspresję genów niezależnie od AGO2 i jego właściwości cięcia transkryptów.

Przeprowadzone przeze mnie eksperymenty wniosły istotny wkład w poznanie mechanizmu wyciszania genów za pomocą art-miRNA A2. Jednocześnie, przy użyciu tego samego układu modelowego, wykonano eksperymenty profilowania polisomów, które wykazały przesunięcie analizowanej puli zmutowanych transkryptów po dodaniu A2 z frakcji ciężkich polisomów, do frakcji zawierających pojedyncze podjednostki rybosomów oraz monosomy. Sugerowało to zahamowanie translacji na etapie inicjacji i/lub wczesnej elongacji. Podsumowując, art-miRNA celujące w wydłużone ciągi powtórzeń CAG znajdujące się w ORF powodują inhibicję translacji oraz gwałtowną deadenylację mRNA, a na późniejszym etapie zachodzi także degradacja transkryptu. W zaproponowanym modelu działania art-miRNA (Rycina 6 w publikacji) kompleksy miRISC nakierowane na wydłużony ciąg powtórzeń znajdują się w bliskiej odległości do ogonów poli(A) badanych transkryptów, co może prowadzić do deadenylacji ogonów poli(A), usunięcia czapeczki 5' i finalnie do degradacji transkryptu. Ponadto, przeprowadzone eksperymenty udowodniły, że art-miRNA A2 może pełnić swoją funkcję wyciszania ekspresji genów niezależnie od AGO2. Do poprawnego działania A2 nie jest wymagana ani właściwość cięcia transkryptu przez AGO2, ani nawet jego obecność w komórce. Sugeruje to, że pozostałe białka AGO występujące w komórkach (AGO1,

AGO3 oraz AGO4) i odpowiadające im kompleksy miRISC mogą funkcjonalnie zastąpić AGO2 i prowadzić do obniżenia ekspresji zmutowanych transkryptów w odpowiedzi na działanie A2.

Poza opisanymi eksperymentami przy których miałem swój udział, druga część omawianej publikacji dotyczyła zrozumienia dlaczego art-miRNA A2, mimo że celuje w wydłużone ciągi CAG, wykazuje zmienną efektywność działania w zależności od modelu choroby poliQ. Wpływ na to mogą mieć różne czynniki komórkowe, jak i cechy charakterystyczne konkretnych transkryptów poliQ. Wśród czynników komórkowych potwierdzono niski poziom ekspresji 4 genów (*HTT*, *ATN1*, *ATXN3* oraz *ATXN7*) w liniach fibroblastów wyprowadzonych od pacjentów, co może przekładać się na zmienną efektywność A2. Przykładowo, gen *ATN1* jest wyciszany z wyższą efektywnością niż pozostałe geny co można tłumaczyć jego wyższym poziomem ekspresji niż pozostałej trójki. Inną poruszoną kwestią była lokalizacja ciągu powtórzeń. Dla allelo-selektywnego działania A2, musi on znajdować się w ORF, gdyż kiedy wydłużony ciąg znajdował się w 3'UTR to wyciszany był zarówno allel zmutowany, jak i normalny (Rycina 2C w publikacji). Powodem prawdopodobnie jest brak translacji 3'UTR. Także specyficzna sekwencja otaczająca ciągi powtórzeń może mieć wpływ na efektywność działania A2 – ma ona m.in. wpływ na stabilność struktur drugorzędowych, które formują się w wyniku występowania wydłużonych ciągów powtórzeń [100]. W niniejszej publikacji potwierdziliśmy także skuteczność testowanej strategii na modelu ludzkich progenitorów neuronalnych HD. Opublikowane wyniki poszerzyły wiedzę odnośnie charakterystyki oraz specyfiki działania art-miRNA, a także pozwoliły na lepsze poznanie mechanizmów działania naturalnych miRNA wiążących się w ORF. Ponadto, badania te podkreślają wszechstronność procesu RNAi oraz dają także nowe możliwości w rozwoju terapii chorób poliQ. Strategia rozwijana jest dalej w Instytucie i testowana na modelach mysich.

6.2. Implications of Poly(A) Tail Processing in Repeat Expansion Diseases

***P. Joachimiak**, A. Ciesiołka, G. Figura, A. Fiszer. Cells. 2022 Feb 15;11(4):677. doi: 10.3390/cells11040677.*

Aby nowo powstałe mRNA mogło pełnić swoje funkcje i przekazywać informację genetyczną, musi przejść proces dojrzewania. Jednym z jego etapów jest proces cięcia i poliadenylacji, czyli dodawania do końca 3' transkryptu adeniny, w wyniku czego powstaje ogon poli(A) [101, 102]. Standardowo zachodzi on w 3'UTR dzięki współwystępowaniu i kooperacji specyficznych czynników *cis* oraz *trans* (z których najistotniejszy to sygnał poliadenylacji, a jego najbardziej kanoniczna sekwencja to AAUAAA), które wspólnie wyznaczają miejsce cięcia transkryptu, a następnie dodawanie niezależnie od matrycy reszt A. APA zachodzi w miejscach występowania niekanonicznych, słabszych sygnałów poliadenylacji i może mieć miejsce nie tylko w 3'UTR, ale także w intronach lub w sekwencjach kodujących transkryptów. Obecnie uważa się, że APA występuje dla 70% genów kodujących białka, a także dla niekodujących RNA, jak długie niekodujące RNA (ang. *long noncoding RNA*, lncRNA) [103–105]. Zaburzenia APA są powiązane się z wieloma chorobami, przede wszystkim z różnego rodzaju nowotworami. Wydaje się, że zagadnienia dotyczące procesów związanych z ogonami poli(A) mogą być istotne dla pełnego opisania zmian molekularnych zachodzących w wielu innych schorzeniach. W omawianej publikacji podsumowałem obecny stan wiedzy odnośnie zaburzeń APA w chorobach spowodowanych ekspansją powtórzeń tandemowych, w tym w chorobach poliQ oraz wskazałem wątki do dalszych badań.

Poza krótkim opisem poliadenylacji i APA, w niniejszej publikacji opisałem kilka metod i narzędzi pozwalających na predykcję i identyfikację miejsc APA oraz bazy danych zbierające informacje nt. już zidentyfikowanych miejsc APA. Są to przydatne narzędzia, w których można wyszukiwać konkretne geny (ludzkie, ale często także mysie i inne) i sprawdzać gdzie APA może wystąpić. Ponadto, bazując na swoich wcześniejszych badaniach, przedstawiłem kilka metod pozwalających na analizę długości ogonów poli(A), co może być szczególnie przydatne w badaniach procesu deadenylacji lub mechanizmów wyciszania ekspresji genów. Opisałem metody do analizy długości ogonów poli(A) konkretnych genów zaczynając od metod historycznych, jak np. metoda

z wykorzystaniem RNazy H/Oligo(dT) i hybrydyzacji typu *northern* [106], a kończąc na używanej przeze mnie metodzie *G/I tailing* [107]. Ponieważ w ostatnich latach częściej w literaturze można spotkać globalne analizy długości ogonów poli(A), toteż krótko nawiązałem do metod takich jak TAIL-Seq [108], PAL-Seq [109] oraz najnowszy z wymienionych - FLAM-Seq [110].

W kolejnej części publikacji skupiłem się na właściwym zagadnieniu, a więc na implikacjach APA w chorobach spowodowanych ekspansją powtórzeń tandemowych, do których zaliczają się choroby poliQ. Grupa ta obejmuje ponad 40 różnych chorób i wraz z rozwojem biologii i genetyki liczba ta stale rośnie [5, 111]. Mutacje w genach odpowiedzialnych za w/w choroby polegają na wydłużeniu ciągu powtórzeń, na który składają się motywy o długości 3-12 nukleotydów [5]. Ponieważ choroby te dotykają przeważnie układu nerwowego oraz mięśniowego, to w publikacji zebrałem dostępne informacje dotyczące procesu poliadenylacji oraz długości ogonów poli(A) w neuronach i mięśniach. Co ciekawe, mRNA w neuronach charakteryzują się wyjątkowo długimi 3'UTRami, a więc dalsze sygnały poliadenylacji (dystalne) są preferencyjnie wybierane w tych komórkach [112–114]. Natomiast inne komórki układu nerwowego jak astrocyty czy oligodendrocyty nie wykazują podobnych tendencji [115, 116]. Komórki mięśniowe także wykazują charakterystyczne wzory poliadenylacji [117], co sugeruje, że wybór konkretnych miejsc poliadenylacji jest tkankowo-specyficzny. Następnie opisałem kilka chorób m.in. HD oraz dystrofię miotoniczną typu 1 (ang. *myotonic dystrophy type 1*, DM1), w których potwierdzono, że występowanie APA ma wpływ na patogenezę.

Z racji tematyki mojej pracy doktorskiej, zależało mi by w opisywanej publikacji szerzej nawiązać do chorób poliQ, w kontekście których rola APA jest najlepiej opisana dla transkryptu *HTT*. Potwierdzono występowanie trzech izoform pełnej długości *HTT* niezwiązanych z mutacją, oraz związanej z mutacją formy *HTT1a* powstającej w wyniku współdziałania zaburzonego splicingu oraz APA. Najprawdopodobniej to wydłużony ciąg powtórzeń CAG w *HTT* może zaburzać splicing, a wraz ze stosunkowo blisko występującym w sekwencji alternatywnym sygnałem poliadenylacji, mogą wpływać na powstawanie takich skróconych, niekompletnych transkryptów. Był to dla mnie punkt wyjścia, aby przeszukać bazy danych APA i sprawdzić czy analogiczne procesy i efekty można obserwować w innych transkryptach poliQ. Postanowiłem skorzystać z jednej

z opisanych w tekście publikacji baz danych zbierających informacje dot. miejsc APA – PolyASite 2.0 [118] – i przeanalizować transkrypty związane z pozostałymi chorobami poliQ pod kątem zidentyfikowanych, potencjalnych miejsc APA. Były dwa kryteria którymi się kierowałem. Po pierwsze, szukałem takich miejsc APA, które stosunkowo często (powyżej wartości 0.1 TPM) były obserwowane w sekwencjonowaniach. Po drugie, biorąc za przykład HD, przeszukałem wspomnianą bazę pod kątem potencjalnych miejsc APA, które spełniały przynajmniej jeden z następujących wymogów: (I) występowały w tym samym egzonie co ciąg powtórzeń CAG; (II) występowały w innym egzonie niż ciąg powtórzeń, ale maksymalnie 500 pz od ciągu powtórzeń (w dojrzałym mRNA); (III) występowały w intronach sąsiadujących z egzonem, w którym znajduje się ciąg powtórzeń (Rycina 2 w publikacji). Szczególną moją uwagę przykuły 3 transkrypty: *ATN1*, *ATXN1* oraz *ATXN3*. *ATN1* oraz *ATXN1* mają wiele potencjalnych miejsc APA w intronach sąsiadujących z egzonem posiadającym ciąg powtórzeń, natomiast w *ATXN3* ciekawym wątkiem wydaje się alternatywnej długości 3'UTR – w stosunku do długości CDS transkryptu jest to bardzo długi 3'UTR, w którym występuje przynajmniej 5 miejsc APA. Występowanie alternatywnej długości 3'UTR w transkryptach może pośrednio wpływać na poziom danego transkryptu – im krótszy 3'UTR, tym np. mniej miejsc wiązania miRNA regulujących ekspresję genu na poziomie mRNA. Aby zobrazować jak alternatywnej długości regiony 3'UTR transkryptów *ATXN3* oraz *HTT* mogą wpływać na ich regulację przez miRNA, przygotowałem odpowiedni schemat (Rycina 3 w publikacji). Widać na nim wyraźnie jak wybór innych miejsc poliadenylacji niż dystalne, prowadzi do utraty miejsc wiązania wielu miRNA, szczególnie w transkrypcie *ATXN3*.

Publikację zakończyłem rozdziałem podsumowującym oraz potencjalnymi perspektywami dalszych badań zagadnienia APA w kontekście chorób spowodowanych ekspansją powtórzeń tandemowych, ze szczególnym naciskiem na choroby poliQ. Poruszyłem aspekt metodologiczny, a mianowicie jak ważny jest dalszy rozwój metod eksperymentalnych i bioinformatycznych prowadzących do identyfikacji miejsc APA. Ponadto, nawiązałem do patogenezy wspomnianych chorób wraz z podkreśleniem, że dotyczą one konkretnych komórek i tkanek, a więc istotnym jest dalsze badanie procesów APA właśnie w tych tkankach i komórkach objętych chorobą. Niemniej

interesująca wydaje się także możliwość wprowadzenia nowych terapii nacelowanych w APA – np. dzięki użyciu ASO można manipulować splicingiem, a także poliadenylacją, co może skutkować powstawaniem transkryptów o konkretnej długości 3'UTR (np. dla *ATXN3*), lub prowadzić do niepowstawania wysoce toksycznych, niekompletnych transkryptów (np. dla *HTT*).

6.3. Precise and accurate allele-specific quantitation of ATXN3 and HTT transcripts in polyQ disease models

***P. Joachimiak**, A. Ciesiołka, E. Kozłowska, P. M. Świtoński, G. Figura, A. Ciołak, G. Adamek, M. Surdyka, Ż. Kalinowska-Pośka, M. Figiel, N. S. Caron, M. R. Hayden, A. Fiszer. BMC Biol. 2023 Feb 1;21(1):17. doi: 10.1186/s12915-023-01515-3.*

W związku z kluczową rolą zmutowanego białka w patogenezie chorób poliQ, powstało wiele publikacji poruszających temat analizy ilościowej zmutowanych białek w tych chorobach [119–122]. Natomiast w przypadku transkryptów poliQ ich dokładna rola nie jest jeszcze znana, choć ukazuje się coraz więcej informacji na ten temat [123–127]. Ponadto, wykazano że zmutowane transkrypty poliQ są obiecującymi celami strategii terapeutycznych zaplanowanych na obniżenie ekspresji zmutowanego genu [13, 22, 128]. Dlatego też kluczowym wydaje się opracowanie niezawodnego podejścia do ilościowego, allelo-selektywnego określenia poziomu badanych alleli, zwłaszcza w komórkach i tkankach nerwowych, które głównie dotknięte są procesami patogennymi prowadzącymi do chorób poliQ.

Ilościowe analizowanie endogennych transkryptów związanych z chorobami poliQ jest jednak skomplikowane. Po pierwsze, ich geny charakteryzują się niskim poziomem ekspresji [129], co może skutkować np. powstawaniem niespecyficznych produktów w reakcjach PCR – amplifikowane mogą być wówczas rejony o zbliżonej sekwencji, ale występujące w znacznie większej liczbie na komórkę (np. sekwencje rRNA). Po drugie, najbardziej oczywistą cechą tych transkryptów pozwalającą na odróżnienie transkryptu WT od MUT jest długość ciągu CAG, jednakże ciągi te występują stosunkowo często w ludzkim genomie/transkryptomie [130], stąd też analiza ciągu w konkretnym genie może być utrudniona. Co istotniejsze jednak, bardzo trudnym (a czasami wręcz niemożliwym) zadaniem jest zaprojektowanie starterów lub sond na ciąg powtórzeń lub na rejon je otaczający. Wynika to m.in. ze „ślizgania się" polimerazy PCR na długich ciągach powtórzeń, występowania struktur II-rzędowych lub powstawania produktów zbyt długich niż jest to zalecane dla qPCR. W związku z tym, postanowiłem zidentyfikować w wybranych ludzkich liniach komórkowych (wyprowadzonych od pacjentów SCA3 i HD) heterozygotyczne warianty SNP występujące w sekwencjach *ATXN3* i *HTT* oraz przypisać je do allelu WT lub MUT. Następnie, wykorzystałem metodę

ddPCR do ilościowego określenia stosunku procentowego alleli WT/MUT w tych liniach komórkowych oraz w bi-allelicznym modelu mysim Hu128/21 [131]. Metoda ddPCR, ze względu na generowanie nawet 20 000 kropel (*droplets*), w których zachodzą niezależne reakcje, jest jeszcze dokładniejsza i bardziej powtarzalna niż qPCR. Ze względu na swoją charakterystykę i odpowiednie zaprojektowanie sond TaqMan, pozwala na bardzo precyzyjne ilościowanie dwóch sekwencji różniących się raptem jednym nukleotydem (jak w przypadku występowania SNP). Ponadto, opisane podejście pozwoliło mi na oszacowanie liczby transkryptów *ATXN3* i *HTT* występujących w pojedynczej komórce, oraz na zaproponowanie użycia opisanej metody do oceny efektywności allelo-selektywnych strategii terapeutycznych w terapii chorób poliQ.

Materiał SCA3 użyty w niniejszej pracy stanowiły linie komórkowe wyprowadzone od pacjenta: fibroblasty, indukowane pluripotentne komórki macierzyste (ang. *induced pluripotent stem cells*, iPSC), neuralne komórki macierzyste (ang. *neural stem cells*, NSC) oraz neurony. Wykorzystana linia fibroblastów SCA3 jest komercyjnie dostępna i została reprogramowana do iPSC w Zakładzie [132] oraz różnicowana do komórek neuronalnych. Taki zestaw linii komórkowych jest bardzo dobrym materiałem do zaplanowanych badań, ponieważ pochodzi od pacjenta, a także pozwala prześledzić cały proces różnicowania neuronalnego komórek, który może mieć wpływ na przedmiot badań. W sekwencji *ATXN3* wspomnianych komórek występuje 5 SNP pozwalających na odróżnienie obu alleli (Rycina 1B w publikacji). Wybrałem dwa z nich, *ATXN3*_SNP2 oraz *ATXN3*_SNP5, aby zaprojektować reakcje ddPCR, które wykorzystywałem w dalszych analizach. *ATXN3*_SNP2 zlokalizowany jest w egzonie 8 *ATXN3*, natomiast *ATXN3*_SNP5 znajduje się w 3'UTR transkryptu. Dodatkowo, są to warianty SNP stosunkowo dobrze opisane w literaturze, gdzie wykazano m.in. że obecność SNP5 powoduje utratę miejsca wiązania konkretnych miRNA [133], co może wiązać się ze zmianą ekspresji *ATXN3*. We wszystkich analizowanych liniach komórkowych otrzymałem zbliżone średnie stosunki procentowe alleli WT/MUT – ok. 54% WT/46% MUT (Ryciny 2A-D w publikacji). Analiza statystyczna przeprowadzona w oparciu o otrzymane wyniki wykazała, że uzyskane stosunki procentowe w różnych typach komórek są do siebie zbliżone, toteż typ komórek nie ma wpływu na stosunek WT/MUT. Tak więc różnicowanie neuronalne w żaden sposób nie zmienia tej wartości dla przebadanego materiału komórkowego

(Rycina 2E w publikacji). Różnice pojawiły się za to, gdy podjąłem się oszacowania łącznej liczby transkryptów (WT+MUT) przypadających na jedną komórkę. Z szacunków wynika, że w fibroblastach i iPSC znajduje się zbliżona liczba transkryptów – odpowiednio 12 i 10 transkryptów *ATXN3* na komórkę. Wynik w fibroblastach jest potwierdzeniem wyników uzyskanych wcześniej w Zakładzie z wykorzystaniem metody smFISH [134]. Natomiast analizy, które przeprowadziłem na materiale uzyskanym z NSC wykazały obecność ok. 54 transkryptów *ATXN3* (WT+MUT) na komórkę (Rycina 2F w publikacji). Wynik ten sugeruje, że różnicowanie neuronalne wpływa z kolei na wzrost liczby transkryptów *ATXN3* na komórkę.

W przypadku HD, badany materiał stanowiły linie komórkowe iPSC, NSC oraz neurony, pochodzące z linii wyprowadzonej od pacjenta z HD. Tym razem w sekwencji *HTT* znajduje się 7 SNP pozwalających na odróżnienie obu alleli (Schemat 3B w publikacji). Do zaprojektowania reakcji ddPCR wybrałem *HTT*_SNP2, *HTT*_SNP5 oraz *HTT*_SNP7, wśród których *HTT*_SNP2 znajduje się w egzonie 50 *HTT*, a *HTT*_SNP5 i *HTT*_SNP7 znajdują się w 3'UTR transkryptu. Są to SNP bardzo dobrze opisane w literaturze, np. *HTT*_SNP7 jest jedynym SNP skorelowanym z występowaniem mutacji *HTT* [9], a wraz z *HTT*_SNP2 były obiektem badań w kontekście allelo-selektywnych terapii z wykorzystaniem ASO prowadzonych przez Wave Life Sciences. W liniach HD otrzymałem następujące stosunki procentowe WT/MUT: 68.7% WT/31.3% MUT w iPSC, 64% WT/36% MUT w NSC oraz 62.4% WT/37.6% MUT w neuronach (Ryciny 4A-C w publikacji). W porównaniu do linii komórkowych SCA3, otrzymałem zatem większą różnicę w poziomie ekspresji obu alleli. Ponadto, analiza statystyczna przeprowadzona w oparciu o uzyskane wyniki pokazała, że różnice w stosunkach WT/MUT pomiędzy komórkami HD są istotne statystycznie (Rycina 4D w publikacji). Tak więc można stwierdzić, że różnicowanie neuronalne wpłynęło na zmianę stosunku ekspresji obu alleli – wraz z nim spadła procentowa ekspresja allelu WT, podczas gdy procentowa ekspresja allelu MUT wzrosła. W komórkach HD także sprawdziłem łączną ilość transkryptów *HTT* przypadających na jedną komórkę. Tym razem obserwowany trend był odwrotny niż w przypadku linii SCA3 - dla iPSC uzyskałem wynik ok. 37 transkryptów *HTT* (WT+MUT) na komórkę, a w NSC było to 16 transkryptów *HTT* (Rycina 4E w publikacji). Wynik

uzyskany dla NSC pokrywał się z wynikami analiz mikroskopowych jakie uzyskaliśmy w Zakładzie w poprzednich badaniach [134].

Kolejnym krokiem było przeanalizowanie, w analogiczny sposób, materiału pochodzącego z mysiego modelu HD. Tym samym, chciałem sprawdzić jak wygląda ekspresja obu alleli w całej tkance, a nie tylko w konkretnej puli komórek. Dodatkowo, chciałem porównać ze sobą różne rejony mózgu i sprawdzić czy istnieją pomiędzy nimi jakieś różnice, a mianowicie czy stosunek WT/MUT różni się pomiędzy prążkowiem (rejonem mózgu najbardziej objętym neurodegeneracją w HD), a innymi rejonami mózgu. W związku z charakterem moich badań konieczne było znalezienie takiego modelu, który posiadałby heterozygotyczne warianty SNP – najlepiej te, które analizowałem w liniach komórkowych HD. Dlatego też skorzystałem z modelu biallelicznego Hu128/21 powstałego przez skrzyżowanie myszy BAC21 oraz YAC128 [131]. Te badania były prowadzone w ramach współpracy z prof. Michaelem Haydenem (University of British Columbia), w którego zespole został otrzymany model Hu128/21, oraz prof. Maciejem Figlem (ICHB PAN), który prowadzi badania z wykorzystaniem tego modelu w Poznaniu. Myszy Hu128/21 są heterozygotami pod względem 2 z 3 wykorzystywanych SNP - *HTT*_SNP2 oraz *HTT*_SNP5. Ze względu na to, że zwierzęta transgeniczne mogą posiadać nierówną liczbę wprowadzonych transgenów (w tym przypadku dotyczy to *HTT* z ciągiem normalnej długości pochodzącym z myszy BAC21 oraz *HTT* z ciągiem zmutowanym pochodzącym z myszy YAC128), w pierwszej kolejności przeanalizowałem gDNA wyizolowane z badanych myszy. Okazało się, że badany model faktycznie posiada nierówną liczbę transgenów – w gDNA otrzymałem stosunek procentowy 64.5% WT/35.5% MUT (Rycina 7A w publikacji), co odbiegało od stosunku 50/50 otrzymanego w obu liniach komórkowych (Ryciny 1D oraz 3D w publikacji). W analizie liczby kopii, po odniesieniu uzyskanych wyników dla transgenu *HTT* do referencyjnego genu *Rpp30* wyniki wskazują, że badane myszy posiadają 11 kopii transgenu pochodzącego z myszy BAC21 oraz 6 kopii transgenu pochodzącego z myszy YAC128 (Rycina 7B w publikacji), co nie zostało wcześniej tak precyzyjnie określone w badaniach z wykorzystaniem tego modelu. Mając na uwadze wyniki uzyskane na gDNA, przeszedłem do analizy ekspresji badanych transgenów. W tym celu, przeanalizowałem 4 rejony mózgu myszy (prążkowie – STR; korę mózgu – CTX;

śródmózgowie – MB; oraz opuszkę węchową – OF) pobrane od myszy 4-miesięcznych i 10-miesięcznych. W przypadku myszy 4-miesięcznych żaden rejon mózgu nie różnił się znacząco pod względem stosunku procentowego WT/MUT od pozostałych (Rycina 7C w publikacji). Natomiast w przypadku myszy 10-miesięcznych zaobserwowałem znaczącą statystycznie różnicę pochodzącą z OF w porównaniu do pozostałych rejonów (Rycina 7D w publikacji). Niemniej jednak, wszystkie uzyskane stosunki procentowe WT/MUT, zarówno u myszy 4- jak i 10-miesięcznych, były zbliżone do wartości uzyskanych podczas analizy gDNA sugerując, że w badanym modelu mysim ekspresja transgenów wynika bezpośrednio z ich liczby kopii. Chcąc sprawdzić, czy wiek myszy wpływa w jakikolwiek sposób na ekspresję transgenów, przeprowadziłem analizę statystyczną w oparciu o w/w wyniki. Kiedy przeanalizowałem wszystkie 4 rejony mózgu łącznie, okazało się, że poziom ekspresji transgenu WT *HTT* był wyższy u myszy 10-miesięcznych, niż u myszy 4-miesięcznych (co za tym idzie, poziom ekspresji MUT *HTT* był niższy). Obserwacja ta była zgodna z innymi badaniami przeprowadzonymi na materiale pobranym od pacjentów HD [135].

Opracowana metoda ddPCR jest bardzo dokładna, powtarzalna i precyzyjna. Wyniki uzyskane z użyciem ddPCR porównałem z odczytami uzyskanymi z RNA-Seq komórek HD iPSC oraz NSC (Rycina S5 w publikacji). Odczyty RNA-Seq były zbliżone do wyników uzyskanych z wykorzystaniem ddPCR, jednakże wartości uzyskane dla pojedynczych powtórzeń biologicznych cechowały się dużo większym rozrzutem, co skutkowało znacznie większymi słupkami błędów. Można z tego wysnuć wniosek, że metoda ddPCR jest dokładniejsza i skuteczniejsza do zaproponowanego podejścia wykorzystującego SNP. Przedstawione podejście eksperymentalne może mieć zastosowanie nie tylko przy bezpośredniej analizie ekspresji dwóch alleli – w publikacji zaproponowałem także inne wykorzystanie metody. Jednym z przykładów jest weryfikacja kontrolnych linii izogenicznych otrzymanych z linii komórkowych HD, tj. linii w których dokonano korekty allelu zmutowanego *HTT* (allel dalej opisywany jako „corrMUT"). W przedstawianej pracy przeanalizowałem dwie linie izogeniczne iPSC (C39 oraz C105) powstałe dzięki zastosowaniu metody CRISPR-Cas9 w komórkach iPSC HD [136]. W obu badanych kontrolnych liniach zaobserwowałem zmienione stosunki procentowe względem wyjściowej linii HD – linia C39 cechowała się stosunkiem 75.9 WT/24.1% corrMUT,

podczas gdy w linii C105 uzyskałem 99.5 WT/0.5% corrMUT (Ryciny 5B,C w publikacji). Pokazuje to, że opisywana metoda może precyzyjnie określić poziom ekspresji obu alleli w modyfikowanych genetycznie liniach kontrolnych, podczas gdy w przypadku qPCR istnieje ryzyko, że różnice będą na tyle subtelne, że nie zostaną dostrzeżone (1 cykl różnicy w qPCR mówi o dwukrotnej różnicy w ekspresji badanego genu - natomiast jeśli różnica wynosi 20%, to otrzymane w qPCR wartości Ct mogą być „zatarte" przez np. błąd pipetowania). Kolejnym potencjalnym wykorzystaniem opisywanego podejścia może być ocena efektywności allelo-selektywnych terapii projektowanych dla chorób poliQ. Pokazują to ryciny 6A i 6B z publikacji, gdzie przedstawiłem działanie oligonukleotydów wyciszających – działającego allelo-selektywnie A2 i nieallelo-selektywnego siHTT. Na rycinie 6A przedstawiłem spadek ekspresji całej *HTT* (WT+MUT), podczas gdy rycina 6B pokazuje spadek ekspresji konkretnego allelu. Takie przedstawienie wyników wyciszenia ekspresji *HTT* z użyciem oligonukleotydów allelo-selektywnych daje więcej dokładniejszych informacji niż wykorzystanie qPCR.

Podsumowując, opisywane podejście wykorzystujące występujące w sekwencji transkryptów SNP, wraz z bardzo dokładną metodą jaką jest ddPCR, pozwala precyzyjnie i powtarzalnie określić poziom ekspresji endogennych alleli WT i MUT, co w przypadku chorób poliQ jest szczególnie istotne i pożądane. Dzięki niemu byłem w stanie wykazać nawet niewielkie różnice pomiędzy badanym materiałem, zarówno komórkowym jak i mysim. Doprowadziło to do postawienia wniosku, że różnicowanie neuronalne może wpływać zarówno na poziom ekspresji konkretnego allelu wyrażonego stosunkiem WT/MUT, jak i na poziom ekspresji w ogóle, wyrażonego szacunkiem liczby transkryptów przypadających na jedną komórkę. Ponadto, przeprowadzone przeze mnie eksperymenty wykazały, że wiek myszy może wpływać na obserwowany poziom ekspresji obu alleli *HTT*. Opisane podejście eksperymentalne jest dokładniejsze od innych, często stosowanych metod i analiz, jak np. qPCR czy RNA-Seq. W publikacji opisałem kilka potencjalnych możliwości i aplikacji do jakich wspomniane podejście można zastosować, takich jak analiza efektywności allelo-selektywnych terapii chorób poliQ oraz sprawdzenie ekspresji genów po zastosowaniu metod inżynierii genetycznej. Jest to istotny i ciekawy wkład do tematyki chorób poliQ.

7. Podsumowanie i dyskusja

Najważniejszymi osiągnięciami i wnioskami opisanymi w rozprawie są:

1. Wkład w kontekst terapeutyczny chorób poliQ przez zaproponowanie mechanizmu wyciszania ekspresji zmutowanych genów poliQ z użyciem art-miRNA, poprzez:

- udział w opracowaniu egzogennego systemu do ilościowej oceny efektów działania oligonukleotydów wyciszających na poziomie celowanego transkryptu oraz powstającego na jego matrycy białka;

- ustalenie, że zahamowanie translacji zmutowanego transkryptu po dodaniu A2 art-miRNA zachodzi wcześniej niż degradacja transkryptu;

- podkreślenie roli procesu deadenylacji w badanym mechanizmie, który prowadzi do degradacji transkryptu w późniejszych punktach czasowych;

- ustalenie sposobu działania A2 art-miRNA niezależnie od właściwości katalitycznych AGO2, jak i w ogóle od obecności AGO2 w komórce.

2. Wkład w dokładniejsze zrozumienie patogenezy chorób poliQ przez lepsze scharakteryzowanie i poznanie biologii transkryptów poliQ, poprzez:

- zgłębienie tematu alternatywnej poliadenylacji w kontekście chorób poliQ;

- przedstawienie na podstawie baz danych najbardziej prawdopodobnych alternatywnych sygnałów poliadenylacji występujących w transkryptach poliQ, wraz ze zwróceniem uwagi na potencjał tych sygnałów w kontekście patogenezy i terapii;

- opracowanie precyzyjnego podejścia eksperymentalnego do ilościowej analizy endogennych transkryptów poliQ;

- określenie stosunku procentowego alleli WT/MUT ataksyny-3 i huntingtyny w różnych liniach komórkowych oraz w tkankach mózgowych myszy;

- oszacowanie łącznej liczby transkryptów ataksyny-3 i huntingtyny w badanych liniach komórkowych;

- zaproponowanie wpływu różnicowania neuronalnego na stosunek alleli WT/MUT oraz liczbę transkryptów przypadających na jedną komórkę;

- zastosowanie w/w podejścia do oceny efektywności allelo-selektywnych terapii chorób poliQ oraz do analizy funkcjonalności genów poddanych modyfikacjom genetycznym.

W trakcie swojego doktoratu pracowałem nad dwoma aspektami dotyczącymi charakterystyki transkryptów poliQ – terapeutycznym oraz związanym z rolą tychże transkryptów w patogenezie chorób poliQ. Obie te kwestie są równie istotne i wzajemnie się przenikają. Kontekst terapeutyczny jest o tyle ważny ponieważ ciągle, mimo cały czas rosnącej wiedzy odnośnie chorób poliQ oraz testowania różnych podejść terapeutycznych, nie opracowano skutecznej strategii ich leczenia. Owszem, z niektórych badań płyną dobrze zapowiadające się wyniki i duży potencjał, jednakże przed naukowcami stoi jeszcze bardzo dużo wyzwań. Wydaje się, że w przypadku HD jednym z istotniejszych jest dążenie do opracowania takiej terapii, która pozwoli na zachowanie w komórce poziomu ekspresji normalnej huntingtyny wystarczającego do poprawnego funkcjonowania komórek pacjentów. Całkowite wyciszenie ekspresji huntingtyny może mieć wiele skutków ubocznych [137], dlatego różne grupy naukowców starają się opracować terapie allelo-selektywne. Oligonukleotyd art-miRNA A2, którego mechanizm działania chciałem poznać, jest w tym przypadku o tyle interesujący, gdyż celuje w powtórzenia CAG, a więc w miejsce mutacji obecne we wszystkich transkryptach poliQ, tak więc u wszystkich pacjentów z chorobami poliQ. A zatem do jego użycia, czy to na liniach komórkowych, czy też zwierzętach, nie ma potrzeby poszukiwania miejsc lub sekwencji charakterystycznych dla allelu zmutowanego, np. konkretnych wariantów SNP. Stanowi to niewątpliwie dużą zaletę tego typu oligonukleotydów, dlatego też tak istotne było poznanie dokładnego mechanizmu jego działania. Niestety, badany art-miRNA nie działa jednakowo skutecznie dla wszystkich transkryptów poliQ – na jego efektywność wpływ mają różne czynniki takie jak m.in. poziom ekspresji danego genu, lokalizacja ciągu powtórzeń oraz sekwencja otaczająca ciąg. Na szczęście istnieją sposoby na poprawienie efektywności działania art-miRNA, np. poprzez zastosowanie wektorów, z których będą ulegać ekspresji [138] lub poprzez wprowadzenie różnych modyfikacji chemicznych [139]. Badania mechanizmu działania art-miRNA mogą zatem stanowić punkt wyjścia do dalszych badań w kontekście terapeutycznym chorób poliQ. Przedstawione wyniki wraz

z zaproponowanym mechanizmem wyciszania mogą pomóc w lepszym zrozumieniu mechanizmów działania endogennych miRNA, gdyż globalne podejścia do identyfikacji miejsc wiązania miRNA ujawniły liczne miejsca zlokalizowane w ORF [90, 91]. Ponadto, istnieją dowody potwierdzające różnice w mechanizmie wyciszania genów za pomocą miRNA w zależności, czy miejsce wiązania miRNA zlokalizowane jest w ORF czy w 3'UTR [92], na co badane cząsteczki stanowią dobry przykład.

Ważnym pytaniem w kontekście chorób poliQ pozostaje to, czy zaangażowane w choroby transkrypty są jedynie pośrednikami wadliwej informacji genetycznej na sekwencję białka, czy mają także swoją rolę w patogenezie chorób poliQ. Swoje badania na ten temat rozpocząłem od analizy procesów poliadenylacji i alternatywnej poliadenylacji, które były niejako kontynuacją eksperymentów dotyczących deadenylacji z poprzedniego projektu. APA jest procesem powszechnie zachodzącym w ludzkich komórkach i jej występowanie powiązano z wieloma chorobami [140–142], stąd też moje zainteresowanie nim. W kontekście chorób poliQ proces ten najlepiej jest opisany dla transkryptu *HTT* co było dla mnie punktem wyjścia, aby przeszukać bazy danych APA i sprawdzić czy analogiczne procesy i efekty można obserwować w innych transkryptach poliQ. Zebrane przeze mnie informacje warto przeanalizować eksperymentalnie, gdyż istnieje szansa, że kiedy potencjalne miejsce poliadenylacji znajduje się blisko wydłużonego ciągu CAG, to w momencie zaburzonego splicingu powstanie w pełni dojrzały i funkcjonalny transkrypt – tak jak ma to miejsce w *HTT*. Ponadto, transkrypty poliQ przeanalizowałem w kontekście różnych izoform 3'UTR. Jest to aspekt, który również warto zbadać eksperymentalnie, gdyż jak pokazuje przykład *HTT*, różne izoformy mogą dominować w różnych tkankach, co przekłada się na tkankowo-specyficzny poziom ekspresji. Gdyby natomiast okazało się, że któraś izoforma danego transkryptu jest powiązana z mutacją, to byłaby ona dobrym celem terapeutycznym np. dla ASO, którymi można modulować wybór miejsc poliadenylacji.

Większość genów ludzkich występuje w liczbie dwóch kopii, jednakże poziom ekspresji tych alleli często nie jest jednakowy. Jest to temat szczególnie ciekawy dla zrozumienia patofizjologii genetycznych chorób dominujących (takich jak większość chorób poliQ), gdzie pacjenci z reguły posiadają jeden allel normalny i jeden zmutowany. Do tej pory niewiele było wiadomo odnośnie poziomów ekspresji obu alleli nawet w najszerzej

opisywanych SCA3 oraz HD. Wiązało się to m.in. z faktem, że ogólna ekspresja *ATXN3* i *HTT* w komórkach jest względnie niska [129]. Ponadto, zmutowane i normalne formy tych transkryptów najłatwiej jest odróżnić na podstawie długości ciągu powtórzeń CAG, jednakże zaprojektowanie jakiejkolwiek metody eksperymentalnej na ciąg powtórzeń jest utrudnione, a czasem wręcz niemożliwe. Dlatego też w swoich eksperymentach postanowiłem wykorzystać warianty SNP występujące naturalnie w sekwencji transkryptów i pozwalające na odróżnienie transkryptu WT od MUT. Przeprowadzenie badań na liniach komórkowych wyprowadzonych z komórek pacjentów dwóch różnych chorób pozwoliło na znalezienie cech wspólnych jak i różnic pomiędzy transkryptami *ATXN3* i *HTT*. W obu tych przypadkach poziom ekspresji allelu WT był wyższy, jednakże pomiędzy transkryptami *HTT* WT i MUT różnica w poziomie ekspresji była większa, niż w przypadku *ATXN3*. Dodatkowo, zaobserwowałem że różnicowanie neuronalne komórek HD wpłynęło na stosunek procentowy ekspresji WT/MUT, podczas gdy w komórkach SCA3 różnicowanie neuronalne nie dawało takiego efektu. Natomiast w obu chorobach różnicowanie wpłynęło na łączną liczbę badanych transkryptów (WT+MUT) przypadających na jedną komórkę, choć w odwrotny sposób – w komórkach SCA3 liczba transkryptów *ATXN3* wzrosła wraz z różnicowaniem, natomiast w HD liczba transkryptów *HTT* spadła. Jest to temat do dalszej analizy, gdyż poziom zmutowanego i normalnego transkryptu ma wpływ na inicjowanie patogenezy na poziomie molekularnym – w przypadku SCA3 może odgrywać rolę znaczny wzrost ekspresji zmutowanego allelu podczas różnicowania neuronalnego, bez zmiany stosunku alleli WT/MUT. Natomiast w HD, istotną rolę może odgrywać zaburzenie proporcji poziomu ekspresji obu alleli na korzyść allelu MUT. Aby móc przełożyć wyniki z przeanalizowanych linii komórkowych generalnie na SCA3 i HD, potrzeba oczywiście dalszych badań na większej puli materiału pochodzącego od pacjentów. Jednakże już teraz można założyć, że transkrypty poliQ mimo zaangażowania w podobne mechanizmy patofizjologiczne, będą wykazywać pewne różnice. Uważam, że jest to temat wart kontynuacji, a nawet rozszerzenia, np. o pozostałe warianty badanych transkryptów powstałe w wyniku alternatywnego splicingu, w tym o transkrypt *HTT1a*. Dzięki wykorzystaniu mysiego modelu HD byłem w stanie sprawdzić jak zmienia się poziom ekspresji obu alleli *HTT* w różnych rejonach mózgu w zależności od wieku myszy. Okazało się, że różnice w stosunku procentowym ekspresji alleli WT/MUT pomiędzy tkankami w danej grupie

wiekowej jeśli występują, to są niewielkie. Natomiast kiedy porównałem obie grupy wiekowe, czyli pre-symptomatyczne myszy 4-miesięczne i symptomatyczne myszy 10-miesięczne to zauważyłem, że wiek myszy miał w tym przypadku wpływ na obserwowany stosunek WT/MUT. Wraz z wiekiem myszy spadł poziom ekspresji allelu MUT w stosunku do allelu WT, co było zgodne z obserwacjami poczynionymi na materiale pobranym od pacjentów w innym badaniu [135]. Może to być związane z tworzeniem się skupisk RNA wynikających z interakcji pomiędzy transkryptem *HTT* pełnej długości, a formą *HTT1a* [55]. Oczywiście, kolejne eksperymenty są niezbędne by określić czy zaobserwowana zależność jest charakterystyczna jedynie dla użytego modelu. Ponadto, aspekt metodyczny publikacji również okazał się być interesujący. Zaproponowane podejście wykorzystujące warianty SNP wraz z użyciem techniki ddPCR okazało się być bardziej precyzyjne i dokładniejsze niż metoda RNA-Seq. W ramach publikacji zaproponowałem także, że opisywane podejście może być wykorzystane do oceny efektywności allelo-selektywnych terapii chorób poliQ oraz do weryfikacji czy metody inżynierii genetycznej (np. CRISPR-Cas9) nie wpływają na ekspresję modyfikowanych alleli. Są to potencjalne skutki uboczne, które może być trudno dostrzec i przeanalizować standardowymi metodami, np. qPCR. Natomiast weryfikacja efektywności terapii z wykorzystaniem opisanego podejścia jest o tyle warta uwagi, gdyż może znacząco ułatwić cały proces, w którym obecnie w testach klinicznych przeprowadza się ilościowe analizy poziomu białka pobranego z płynu mózgowo-rdzeniowego.

Podsumowując, wyniki które otrzymałem wnoszą wkład zarówno w kontekst patogenezy chorób poliQ, ich potencjalne terapie, a także poruszają ważny aspekt metodyczny. W swoich badaniach chciałem potwierdzić i podkreślić, że transkrypty związane z chorobami poliQ są dobrymi obiektami do analiz dotyczących zarówno molekularnych mechanizmów choroby, jak i planowania kolejnych strategii terapeutycznych. W tym celu korzystałem z różnych modeli – w analizach dot. aspektu terapeutycznego i mechanizmu działania art-miRNA kluczowe było użycie systemu indukowalnej nadekspresji z układem reporterowym. Taki system dawał mi możliwość manipulacji ekspresją egzogenu w zależności od potrzeb i omijał problem niskiej ekspresji endogenu. Z kolei w trakcie badań funkcjonowania transkryptów i patogenezy

chorób poliQ pożądanym celem analiz był endogen, tak aby starać się możliwie jak najdokładniej odzwierciedlić naturalnie występujący system. Oczywiście, jeszcze wiele aspektów patogenezy chorób poliQ pozostaje do rozwikłania. Być może najistotniejszym pytaniem jest dlaczego choroby poliQ, mimo że zmutowane transkrypty i białka są obecne prawie we wszystkich typach komórek, dotykają układu nerwowego i prowadzą do neurodegeneracji. Zwłaszcza, że w zależności od choroby różne typy neuronów są szczególnie narażone na degenerację, a taka selektywność stanowi też trudność przy projektowaniu strategii terapeutycznych. Poznanie odpowiedzi na te pytania jest wyjątkowo ważne, ponieważ nasza niewiedza w tym aspekcie na pewno wpływa na fakt, że choroby poliQ są cały czas nieuleczalne.

8. Wykaz używanych skrótów

poliQ – choroby poliglutaminowe

ORF – otwarta ramka odczytu

SCA – ataksja rdzeniowo-móżdżkowa

HD – choroba Huntingtona

RNAi – interferencja RNA

ASO – oligonukleotydy antysensowe

HTT - huntingtyna

ATXN3 – ataksyna-3

sCAG - małe RNA CAG

APA – alternatywna poliadenylacja

NCL - nukleolina

SNP – polimorfizm pojedynczego nukleotydu

siRNA – małe interferujące RNA

shRNA – krótkie RNA o strukturze spinki do włosów

art-miRNA – sztuczne miRNA celujące w rejon powtórzeń CAG

Fluc – lucyferaza Firefly

NLucP – lucyferaza Nano z domeną PEST

siHTT – siRNA celujące w huntingtynę

siRL – siRNA kontrolny

iPSC – indukowalne pluripotentne komórki macierzyste

NSC – neuralne komórki macierzyste

9. Bibliografia

1. Lieberman AP, Shakkottai VG, Albin RL. Polyglutamine Repeats in Neurodegenerative Diseases. Annu Rev Pathol Mech Dis. 2019;14:1–27.

2. Swami M, Hendricks AE, Gillis T, Massood T, Mysore J, Myers RH, et al. Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. Hum Mol Genet. 2009;18:3039–47.

3. Khristich AN, Mirkin SM. On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. J Biol Chem. 2020;295:4134–70.

4. Stoyas CA, La Spada AR. The CAG–polyglutamine repeat diseases: a clinical, molecular, genetic, and pathophysiologic nosology. In: Handbook of Clinical Neurology. Elsevier; 2018. p. 143–70.

5. Paulson H. Repeat Expansion Diseases. In: Handb Clin Neurol. Springer Berlin Heidelberg; 2018. p. 105–23.

6. Bauer PO, Nukina N. The pathogenic mechanisms of polyglutamine diseases and current therapeutic strategies. Journal of Neurochemistry. 2009;110:1737–65.

7. Keiser MS, Kordasiewicz HB, McBride JL. Gene suppression strategies for dominantly inherited neurodegenerative diseases: lessons from Huntington's disease and spinocerebellar ataxia. Human Molecular Genetics. 2016;25:R35–64.

8. Wild EJ, Tabrizi SJ. Therapies targeting DNA and RNA in Huntington's disease. The Lancet Neurology. 2017;16:837–47.

9. Pfister EL, Kennington L, Straubhaar J, Wagh S, Liu W, DiFiglia M, et al. Five siRNAs Targeting Three SNPs May Provide Therapy for Three-Quarters of Huntington's Disease Patients. Curr Biol. 2009;19:774–8.

10. Scoles DR, Meera P, Schneider MD, Paul S, Dansithong W, Figueroa KP, et al. Antisense oligonucleotide therapy for spinocerebellar ataxia type 2. Nature. 2017;544:362–6.

11. Moore LR, Rajpal G, Dillingham IT, Qutob M, Blumenstein KG, Gattis D, et al. Evaluation of Antisense Oligonucleotides Targeting ATXN3 in SCA3 Mouse Models. Mol Ther - Nucleic Acids. 2017;7 June:200–10.

12. Monteys AM, Ebanks SA, Keiser MS, Davidson BL. CRISPR/Cas9 Editing of the Mutant Huntingtin Allele In Vitro and In Vivo. Mol Ther. 2017;25:12–23.

13. Bunting EL, Hamilton J, Tabrizi SJ. Polyglutamine diseases. Curr Opin Neurobiol. 2022;72:39–47.

14. Nance MA, Myers RH. Juvenile onset Huntington's disease - Clinical and research perspectives. Mental Retardation and Developmental Disabilities Research Reviews. 2001;7:153–7.

15. Ghosh R, Tabrizi SJ. Huntington disease. 1st edition. Elsevier B.V.; 2018.

16. Zeitlin S, Liu JP, Chapman DL, Papaioannou VE, Efstratiadis A. Increased apoptosis and early embryonic lethality in mice nullizygous for the Huntington's disease gene homologue. Nat Genet. 1995;11:155–63.

17. Gauthier LR, Charrin BC, Borrell-Pagès M, Dompierre JP, Rangone H, Cordelières FP, et al. Huntingtin controls neurotrophic support and survival of neurons by enhancing BDNF vesicular transport along microtubules. Cell. 2004;118:127–38.

18. Saudou F, Humbert S. The Biology of Huntingtin. Neuron. 2016;89:910–26.

19. Sathasivam K, Neueder A, Gipson TA, Landles C, Benjamin AC, Bondulich MK, et al. Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. Proc Natl Acad Sci U S A. 2013;110:2366–70.

20. Mangiarini L, Sathasivam K, Seller M, Cozens B, Harper A, Hetherington C, et al. Exon 1 of the HD Gene with an Expanded CAG Repeat Is Sufficient to Cause a Progressive Neurological Phenotype in Transgenic Mice. Cell. 1996;87:493–506.

21. Scherzinger E, Lurz R, Turmaine M, Mangiarini L, Hollenbach B, Hasenbank R, et al. Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. Cell. 1997;90:549–58.

22. Matos CA, de Almeida LP, Nóbrega C. Machado–Joseph disease/spinocerebellar ataxia type 3: lessons from disease pathogenesis and clues into therapy. Journal of Neurochemistry. 2019;148:8–28.

23. Burnett B, Li F, Pittman RN. The polyglutamine neurodegenerative protein ataxin-3 binds polyubiquitylated proteins and has ubiquitin protease activity. Hum Mol Genet. 2003;12:3195–205.

24. Herzog LK, Kevei É, Marchante R, Böttcher C, Bindesbøll C, Lystad AH, et al. The Machado–Joseph disease deubiquitylase ataxin-3 interacts with LC3C/GABARAP and promotes autophagy. Aging Cell. 2020;19.

25. Pohl C, Dikic I. Cellular quality control by the ubiquitin-proteasome system and autophagy. Science (80-). 2019;366:818–22.

26. Evert BO, Vogt IR, Vieira-Saecker AM, Ozimek L, De Vos RAI, Brunt ERP, et al. Gene expression profiling in ataxin-3 expressing cell lines reveals distinct effects of normal and mutant ataxin-3. J Neuropathol Exp Neurol. 2003;62:1006–18.

27. Gao R, Chakraborty A, Geater C, Pradhan S, Gordon KL, Snowden J, et al. Mutant huntingtin impairs PNKP and ATXN3, disrupting DNA repair and transcription. Elife. 2019;8:1–31.

28. Evers MM, Toonen LJA, Van Roon-Mom WMC. Ataxin-3 protein and RNA toxicity in spinocerebellar ataxia type 3: Current insights and emerging therapeutic strategies. Molecular Neurobiology. 2014;49:1513–31.

29. Nóbrega C, Simões AT, Duarte-Neves J, Duarte S, Vasconcelos-Ferreira A, Cunha-Santos J, et al. Molecular Mechanisms and Cellular Pathways Implicated in Machado-Joseph Disease Pathogenesis. Adv Exp Med Biol. 2018;1049:349–67.

30. Ross CA, Wood JD, Schilling G, Peters MF, Nucifora FC, Cooper JK, et al. Polyglutamine pathogenesis. Philos Trans R Soc B Biol Sci. 1999;354:1005–11.

31. Wanker EE. Protein Aggregation and Pathogenesis of Huntington's Disease: Mechanisms and Correlations. Biol Chem. 2000;381:937–42.

32. Zoghbi HY, Orr HT. Polyglutamine diseases: Protein cleavage and aggregation. Curr Opin Neurobiol. 1999;9:566–70.

33. Sobczak K, de Mezer M, Michlewski G, Krol J, Krzyzosiak WJ. RNA structure of trinucleotide repeats associated with human neurological diseases. Nucleic Acids Res. 2003;31:5469–82.

34. Michalowski S, Miller JW, Urbinati CR, Paliouras M, Swanson MS, Griffith J. Visualization of double-stranded RNAs from the myotonic dystrophy protein kinase gene and interactions with CUG-binding protein. Nucleic Acids Res. 1999;27:3534–42.

35. Tian B, White RJ, Xia T, Welle S, Turner DH, Mathews MB, et al. Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. RNA. 2000;6:79–87.

36. Li LB, Yu Z, Teng X, Bonini NM. RNA toxicity is a component of ataxin-3 degeneration in Drosophila. Nature. 2008;453:1107–11.

37. Ginisty H, Amalric F, Bouvet P. Nucleolin functions in the first step of ribosomal RNA processing. EMBO J. 1998;17:1476–86.

38. Ginisty H, Serin G, Ghisolfi-Nieto L, Roger B, Libante V, Amalric F, et al. Interaction of nucleolin with an evolutionarily conserved pre-ribosomal RNA sequence is required for the assembly of the primary

processing complex. J Biol Chem. 2000;275:18845–50.

39. Abdelmohsen K, Gorospe M. RNA-binding protein nucleolin in disease. RNA Biol. 2012;9:799–808.

40. Jia W, Yao Z, Zhao J, Guan Q, Gao L. New perspectives of physiological and pathological functions of nucleolin (NCL). Life Sci. 2017;186 May:1–10.

41. Qiu W, Zhou F, Zhang Q, Sun X, Shi X, Liang Y, et al. Overexpression of nucleolin and different expression sites both related to the prognosis of gastric cancer. Apmis. 2013;121:919–25.

42. Pichiorri F, Palmieri D, De Luca L, Consiglio J, You J, Rocci A, et al. In vivo NCL targeting affects breast cancer aggressiveness through miRNA regulation. J Exp Med. 2013;210:951–68.

43. Caudle WM, Kitsou E, Li J, Bradner J, Zhang J. A role for a novel protein, nucleolin, in Parkinson's disease. Neurosci Lett. 2009;459:11–5.

44. Parlato R, Kreiner G. Nucleolar activity in neurodegenerative diseases: A missing piece of the puzzle? J Mol Med. 2013;91:541–7.

45. Tsoi H, Lau TCK, Tsang SY, Lau KF, Chan HYE. CAG expansion induces nucleolar stress in polyglutamine diseases. Proc Natl Acad Sci U S A. 2012;109:13428–33.

46. Deisenroth C, Zhang Y. Ribosome biogenesis surveillance: Probing the ribosomal protein-Mdm2-p53 pathway. Oncogene. 2010;29:4253–60.

47. An Y, Chen ZS, Chan HYE, Ngo JCK. Molecular insights into the interaction of CAG trinucleotide RNA repeats with nucleolin and its implication in polyglutamine diseases. Nucleic Acids Res. 2022;50:7655–68.

48. Krol J, Fiszer A, Mykowska A, Sobczak K, de Mezer M, Krzyzosiak WJ. Ribonuclease Dicer Cleaves Triplet Repeat Hairpins into Shorter Repeats that Silence Specific Targets. Mol Cell. 2007;25:575–86.

49. Abolhassani N, Iyama T, Tsuchimoto D, Sakumi K, Ohno M, Behmanesh M, et al. NUDT16 and ITPA play a dual protective role in maintaining chromosome stability and cell growth by eliminating dIDP/IDP and dITP/ITP from nucleotide pools in mammals. Nucleic Acids Res. 2010;38:2891–903.

50. Iyama T, Abolhassani N, Tsuchimoto D, Nonaka M, Nakabeppu Y. NUDT16 is a (deoxy)inosine diphosphatase, and its deficiency induces accumulation of single-strand breaks in nuclear DNA and growth arrest. Nucleic Acids Res. 2010;38:4834–43.

51. Peng S, Guo P, Lin X, An Y, Sze KH, Lau MHY, et al. Cag rnas induce dna damage and apoptosis by silencing nudt16 expression in polyglutamine degeneration. Proc Natl Acad Sci U S A. 2021;118.

52. Bañez-Coronel M, Porta S, Kagerbauer B, Mateu-Huertas E, Pantano L, Ferrer I, et al. A pathogenic mechanism in huntington's disease involves small CAG-repeated RNAs with neurotoxic activity. PLoS Genet. 2012;8.

53. Neueder A, Landles C, Ghosh R, Howland D, Myers RH, Faull RLM, et al. The pathogenic exon 1 HTT protein is produced by incomplete splicing in Huntington's disease patients. Sci Rep. 2017;7.

54. Mason MA, Gomez-Paredes C, Sathasivam K, Neueder A, Papadopoulou AS, Bates GP. Silencing Srsf6 does not modulate incomplete splicing of the huntingtin gene in Huntington's disease models. Sci Rep. 2020;10:14057.

55. Fienko S, Landles C, Sathasivam K, McAteer SJ, Milton RE, Osborne GF, et al. Alternative processing of human HTT mRNA with implications for Huntington's disease therapeutics. Brain. 2022;145:4409–4424.

56. Lin B, Rommens JM, Graham RK, Kalchman M, Macdonald H, Nasir J, et al. Differential 3' polyadenylation of the huntington disease gene results in two mRNA species with variable tissue expression. Hum Mol Genet. 1993;2:1541–5.

57. Romo L, Ashar-Patel A, Pfister E, Aronin N. Alterations in mRNA 3' UTR Isoform Abundance

Accompany Gene Expression Changes in Human Huntington's Disease Brains. Cell Rep. 2017;20:3057–70.

58. Eddings CR, Arbez N, Akimov S, Geva M, Hayden MR, Ross CA. Pridopidine protects neurons from mutant-huntingtin toxicity via the sigma-1 receptor. Neurobiol Dis. 2019;129 September 2018:118–29.

59. Grachev ID, Meyer PM, Becker GA, Bronzel M, Marsteller D, Pastino G, et al. Sigma-1 and dopamine D2/D3 receptor occupancy of pridopidine in healthy volunteers and patients with Huntington disease: a [18F] fluspidine and [18F] fallypride PET study. Eur J Nucl Med Mol Imaging. 2021;48:1103–15.

60. Ravikumar B, Vacher C, Berger Z, Davies JE, Luo S, Oroz LG, et al. Inhibition of mTOR induces autophagy and reduces toxicity of polyglutamine expansions in fly and mouse models of Huntington disease. Nat Genet. 2004;36:585–95.

61. Her LS, Goldstein LSB. Enhanced sensitivity of striatal neurons to axonal transport defects induced by mutant huntingtin. J Neurosci. 2008;28:13662–72.

62. Leavitt BR, Kordasiewicz HB, Schobel SA. Huntingtin-Lowering Therapies for Huntington Disease: A Review of the Evidence of Potential Benefits and Risks. JAMA Neurol. 2020;77:764–72.

63. Tabrizi SJ, Estevez-Fraga C, van Roon-Mom WMC, Flower MD, Scahill RI, Wild EJ, et al. Potential disease-modifying therapies for Huntington's disease: lessons learned and future opportunities. Lancet Neurol. 2022;21:645–58.

64. Tabrizi SJ, Ghosh R, Leavitt BR. Huntingtin Lowering Strategies for Disease Modification in Huntington's Disease. Neuron. 2019;101:801–19.

65. Dhuri K, Bechtold C, Quijano E, Pham H, Gupta A, Vikram A, et al. Antisense oligonucleotides: An emerging area in drug discovery and development. Journal of Clinical Medicine. 2020;9:1–24.

66. Kingwell K. Double setback for ASO trials in Huntington disease. Nature reviews. Drug discovery. 2021;20:412–3.

67. Friedrich J, Kordasiewicz HB, O'Callaghan B, Handler HP, Wagener C, Duvick L, et al. Antisense oligonucleotide-mediated ataxin-1 reduction prolongs survival in SCA1 mice and reveals disease-associated transcriptome profiles. JCI insight. 2018;3.

68. McLoughlin HS, Moore LR, Chopra R, Komlo R, McKenzie M, Blumenstein KG, et al. Oligonucleotide therapy mitigates disease in spinocerebellar ataxia type 3 mice. Ann Neurol. 2018;84:64–77.

69. Toonen LJA, Rigo F, van Attikum H, van Roon-Mom WMC. Antisense Oligonucleotide-Mediated Removal of the Polyglutamine Repeat in Spinocerebellar Ataxia Type 3 Mice. Mol Ther - Nucleic Acids. 2017;8 September:232–42.

70. Niu C, Prakash TP, Kim A, Quach JL, Huryn LA, Yang Y, et al. Antisense oligonucleotides targeting mutant Ataxin-7 restore visual function in a mouse model of spinocerebellar ataxia type 7. Sci Transl Med. 2018;10.

71. Fiszer A, Olejniczak M, Galka-Marciniak P, Mykowska A, Krzyzosiak WJ. Self-duplexing CUG repeats selectively inhibit mutant huntingtin expression. Nucleic Acids Res. 2013;41:10426–37.

72. Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature. 2001;411:494–8.

73. Djuranovic S, Nahvi A, Green R. miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. Science (80-). 2012;336:237–40.

74. Bazzini AA, Lee MT, Giraldez AJ. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in Zebrafish. Science (80-). 2012;336:233–7.

75. Chendrimada TP, Finn KJ, Ji X, Baillat D, Gregory RI, Liebhaber SA, et al. MicroRNA silencing through RISC recruitment of eIF6. Nature. 2007;447:823–8.

76. Maroney PA, Yu Y, Fisher J, Nilsen TW. Evidence that microRNAs are associated with translating messenger RNAs in human cells. Nat Struct Mol Biol. 2006;13:1102–7.

77. Petersen CP, Bordeleau ME, Pelletier J, Sharp PA. Short RNAs repress translation after initiation in mammalian cells. Mol Cell. 2006;21:533–42.

78. Naeli P, Winter T, Hackett AP, Alboushi L, Jafarnejad SM. The intricate balance between microRNA-induced mRNA decay and translational repression. FEBS Journal. 2022.

79. Park JE, Yi H, Kim Y, Chang H, Kim VN. Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. Mol Cell. 2016;62:462–71.

80. Eisen TJ, Eichhorn SW, Subtelny AO, Lin KS, McGeary SE, Gupta S, et al. The Dynamics of Cytoplasmic mRNA Metabolism. Mol Cell. 2020;77:786-799.e10.

81. Zinder JC, Lima CD. Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. Genes Dev. 2017;31:88–100.

82. Braun JE, Truffault V, Boland A, Huntzinger E, Chang C Te, Haas G, et al. A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5' exonucleolytic degradation. Nat Struct Mol Biol. 2012;19:1324–31.

83. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature. 2010;466:835–40.

84. Eichhorn SW, Guo H, McGeary SE, Rodriguez-Mias RA, Shin C, Baek D, et al. MRNA Destabilization Is the dominant effect of mammalian microRNAs by the time substantial repression ensues. Mol Cell. 2014;56:104–15.

85. Béthune J, Artus-Revel CG, Filipowicz W. Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. EMBO Rep. 2012;13:716–23.

86. Larsson O, Nadon R. Re-analysis of genome wide data on mammalian microRNA-mediated suppression of gene expression. Translation. 2013;1:e24557.

87. Duchaine TF, Fabian MR. Mechanistic insights into microrna-mediated gene silencing. Cold Spring Harb Perspect Biol. 2019;11.

88. Biasini A, Abdulkarim B, Pretis S, Tan JY, Arora R, Wischnewski H, et al. Translation is required for miRNA-dependent decay of endogenous transcripts. EMBO J. 2021;40:1–15.

89. Jin HY, Xiao C. MicroRNA mechanisms of action: What have we learned from mice? Frontiers in Genetics. 2015;6 NOV:328.

90. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. Cell. 2010;141:129–41.

91. Zhang K, Zhang X, Cai Z, Zhou J, Cao R, Zhao Y, et al. A novel class of microRNA-recognition elements that function only within open reading frames. Nat Struct Mol Biol. 2018;25:1019–27.

92. Hausser J, Syed AP, Bilen B, Zavolan M. Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. Genome Res. 2013;23:604–15.

93. Shin E, Jin H, Suh D, Luo Y, Ha H, Kim TH, et al. An alternative miRISC targets a cancer-associated coding sequence mutation in FOXL2. EMBO J. 2021;40:1–23.

94. Fiszer A, Mykowska A, Krzyzosiak WJ. Inhibition of mutant huntingtin expression by RNA duplex targeting expanded CAG repeats. Nucleic Acids Res. 2011;39:5578–85.

95. Fiszer A, Wroblewska JP, Nowak BM, Krzyzosiak WJ. Mutant CAG repeats effectively targeted by RNA interference in SCA7 cells. Genes (Basel). 2016;7.

96. Urbanek MO, Fiszer A, Krzyzosiak WJ. Reduction of Huntington's disease RNA Foci by CAG repeat-targeting reagents. Front Cell Neurosci. 2017;11 March:1–13.

97. Hu J, Liu J, Yu D, Chu Y, Corey DR. Mechanism of allele-selective inhibition of huntingtin expression by duplex RNAs that target CAG repeats: Function through the RNAi pathway. Nucleic Acids Res. 2012;40:11270–80.

98. Liu J, Carmell MA, Rivas F V., Marsden CG, Thomson JM, Song JJ, et al. Argonaute2 is the catalytic engine of mammalian RNAi. Science (80-). 2004;305:1437–41.

99. Rivas F V., Tolia NH, Song JJ, Aragon JP, Liu J, Hannon GJ, et al. Purified Argonaute2 and an siRNA form recombinant human RISC. Nat Struct Mol Biol. 2005;12:340–9.

100. Ciesiolka A, Jazurek M, Drazkowska K, Krzyzosiak WJ. Structural characteristics of simple RNA repeats associated with disease and their deleterious protein interactions. Front Cell Neurosci. 2017;11 April:1–19.

101. Casañal A, Kumar A, Hill CH, Easter AD, Emsley P, Degliesposti G, et al. Architecture of eukaryotic mRNA 3'-end processing machinery. Science (80-). 2017;358:1056–9.

102. Passmore LA, Coller J. Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. Nat Rev Mol Cell Biol. 2021. https://doi.org/10.1038/s41580-021-00417-y.

103. Tian B, Manley JL. Alternative polyadenylation of mRNA precursors. Nat Rev Mol Cell Biol. 2016;18:18–30.

104. Elkon R, Ugalde AP, Agami R. Alternative cleavage and polyadenylation: Extent, regulation and function. Nat Rev Genet. 2013;14:496–506.

105. Sommerkamp P, Cabezas-Wallscheid N, Trumpp A. Alternative Polyadenylation in Stem Cell Self-Renewal and Differentiation. Trends Mol Med. 2021;27:660–72.

106. Sippel AE, Stavrianopoulos JG, Schutz G, Feigelson P. Translational properties of rabbit globin mRNA after specific removal of poly(A) with ribonuclease H. Proc Natl Acad Sci U S A. 1974;71:4635–9.

107. Kusov YY, Shatirishvili G, Dzagurov G, Gauss-Müller V. A new G-tailing method for the determination of the poly(A) tail length applied to hepatitis A virus RNA. Nucleic Acids Res. 2001;29:E57-7.

108. Chang H, Lim J, Ha M, Kim VN. TAIL-seq: Genome-wide determination of poly(A) tail length and 3' end modifications. Mol Cell. 2014;53:1044–52.

109. Subtelny AO, Eichhorn SW, Chen GR, Sive H, Bartel DP. Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature. 2014;508:66–71.

110. Legnini I, Alles J, Karaiskos N, Ayoub S, Rajewsky N. FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control. Nat Methods. 2019;16:879–86.

111. Malik I, Kelley CP, Wang ET, Todd PK. Molecular mechanisms underlying nucleotide repeat expansion disorders. Nat Rev Mol Cell Biol. 2021;22:589–607.

112. Smibert P, Miura P, Westholm JO, Shenker S, May G, Duff MO, et al. Global Patterns of Tissue-Specific Alternative Polyadenylation in Drosophila. Cell Rep. 2012;1:277–89.

113. Miura P, Shenker S, Andreu-Agullo C, Westholm JO, Lai EC. Widespread and extensive lengthening of 3'UTRs in the mammalian brain. Genome Res. 2013;23:812–25.

114. Wei L, Lee S, Majumdar S, Zhang B, Sanfilippo P, Joseph B, et al. Overlapping Activities of ELAV/Hu Family RNA Binding Proteins Specify the Extended Neuronal 3' UTR Landscape in Drosophila. Mol Cell. 2020;80:140-155.e6.

115. Wang W, Wei Z, Li H. A change-point model for identifying 3'UTR switching by next-generation RNA sequencing. Bioinformatics. 2014;30:2162–70.

116. Xia Z, Donehower LA, Cooper TA, Neilson JR, Wheeler DA, Wagner EJ, et al. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. Nat

Commun. 2014;5.

117. Blazie SM, Babb C, Wilky H, Rawls A, Park JG, Mangone M. Comparative RNA-Seq analysis reveals pervasive tissue-specific alternative polyadenylation in Caenorhabditis elegans intestine and muscles. BMC Biol. 2015;13.

118. Herrmann CJ, Schmidt R, Kanitz A, Artimo P, Gruber AJ, Zavolan M. PolyASite 2.0: A consolidated atlas of polyadenylation sites from 3' end sequencing. Nucleic Acids Res. 2020;48:D174–9.

119. Baldo B, Sajjad MU, Cheong RY, Bigarreau J, Vijayvargia R, McLean C, et al. Quantification of total and mutant huntingtin protein levels in biospecimens using a novel alphaLISA assay. eNeuro. 2018;5:234–52.

120. Wild EJ, Boggio R, Langbehn D, Robertson N, Haider S, Miller JRC, et al. Quantification of mutant huntingtin protein in cerebrospinal fluid from Huntington's disease patients. J Clin Invest. 2015;125:1979–86.

121. Fodale V, Pintauro R, Daldin M, Altobelli R, Spiezia MC, Bisbocci M, et al. Analysis of mutant and total huntingtin expression in Huntington's disease murine models. Sci Rep. 2020;10.

122. Gonsior K, Kaucher GA, Pelz P, Schumann D, Gansel M, Kuhs S, et al. PolyQ-expanded ataxin-3 protein levels in peripheral blood mononuclear cells correlate with clinical parameters in SCA3: a pilot study. J Neurol. 2021;268:1304–15.

123. Nalavade R, Griesche N, Ryan DP, Hildebrand S, Krauß S. Mechanisms of RNA-induced toxicity in CAG repeat disorders. Cell Death and Disease. 2013;4:752.

124. Chan HYE. RNA-mediated pathogenic mechanisms in polyglutamine diseases and amyotrophic lateral sclerosis. Frontiers in Cellular Neuroscience. 2014;8 DEC.

125. Martí E. RNA toxicity induced by expanded CAG repeats in Huntington's disease. Brain Pathol. 2016;26:779–86.

126. Fiszer A, Krzyzosiak WJ. RNA toxicity in polyglutamine disorders: Concepts, models, and progress of research. Journal of Molecular Medicine. 2013;91:683–91.

127. Ly S, Didiot M-C, Ferguson CM, Coles AH, Miller R, Chase K, et al. Mutant huntingtin messenger RNA forms neuronal nuclear clusters in rodent and human brains. Brain Commun. 2022;4.

128. Fiszer A, Krzyzosiak WJ. Oligonucleotide-based strategies to combat polyglutamine diseases. Nucleic Acids Res. 2014;42:6787–810.

129. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature. 2012;489:391–9.

130. Butland SL, Devon RS, Huang Y, Mead CL, Meynert AM, Neal SJ, et al. CAG-encoded polyglutamine length polymorphism in the human genome. BMC Genomics. 2007;8:1–18.

131. Southwell AL, Skotte NH, Villanueva EB, Østergaard ME, Gu X, Kordasiewicz HB, et al. A novel humanized mouse model of Huntington disease for preclinical development of therapeutics targeting mutant huntingtin alleles. Hum Mol Genet. 2017;26:1115–32.

132. Ciolak A, Krzyzosiak WJ, Kozlowska E, Fiszer A. Generation of human iPS cell line IBCHi002-A from spinocerebellar ataxia type 3/Machado-Joseph disease patient's fibroblasts. Stem Cell Res. 2020;45.

133. Melo ARV, Raposo M, Ventura M, Martins S, Pavão S, Alonso I, et al. Genetic Variation in ATXN3 (Ataxin-3) 3'UTR: Insights into the Downstream Regulatory Elements of the Causative Gene of Machado-Joseph Disease/Spinocerebellar Ataxia Type 3. Cerebellum. 2022. https://doi.org/10.1007/s12311-021-01358-0.

134. Ciesiolka A, Stroynowska-Czerwinska A, Joachimiak P, Ciolak A, Kozlowska E, Michalak M, et al. Artificial miRNAs targeting CAG repeat expansion in ORFs cause rapid deadenylation and translation inhibition of mutant transcripts. Cell Mol Life Sci. 2020. https://doi.org/10.1007/s00018-020-03596-7.

135. Liu W, Chaurette J, Pfister EL, Kennington LA, Chase KO, Bullock J, et al. Increased steady-state mutant huntingtin mrna in huntington's disease brain. J Huntingtons Dis. 2013;2:491–500.

136. Dabrowska M, Ciolak A, Kozlowska E, Fiszer A, Olejniczak M. Generation of new isogenic models of huntington's disease using CRISPR-Cas9 technology. Int J Mol Sci. 2020;21.

137. Liu JP, Zeitlin SO. Is huntingtin dispensable in the adult brain? Journal of Huntington's Disease. 2017;6:1–17.

138. Kotowska-Zimmer A, Ostrovska Y, Olejniczak M. Universal RNAi Triggers for the Specific Inhibition of Mutant Huntingtin, Atrophin-1, Ataxin-3, and Ataxin-7 Expression. Mol Ther - Nucleic Acids. 2020;19 March:562–71.

139. Yu D, Pendergraff H, Liu J, Kordasiewicz HB, Cleveland DW, Swayze EE, et al. Single-stranded RNAs use RNAi to potently and allele-selectively inhibit mutant huntingtin expression. Cell. 2012;150:895–908.

140. Curinha A, Braz SO, Pereira-Castro I, Cruz A, Moreira A. Implications of polyadenylation in health and disease. Nucleus. 2014;5:508–19.

141. Gruber AJ, Zavolan M. Alternative cleavage and polyadenylation in health and disease. Nat Rev Genet. 2019;20:599–614.

142. Patel R, Brophy C, Hickling M, Neve J, Furger A. Alternative cleavage and polyadenylation of genes associated with protein turnover and mitochondrial function are deregulated in Parkinson's, Alzheimer's and ALS disease. BMC Med Genomics. 2019;12:1–14.

Oświadczenia określające wkład doktoranta
w powstanie publikacji naukowych wchodzących
w skład rozprawy doktorskiej

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, 13.03.2023r.

Paweł Joachimiak
Zakład Biotechnologii Medycznej
Instytut Chemii Bioorganicznej PAN
ul. Noskowskiego 12/14, Poznań

OŚWIADCZENIE KANDYDATA O WŁASNYM WKŁADZIE W PUBLIKACJE NAUKOWE WCHODZĄCE W SKŁAD ROZPRAWY DOKTORSKIEJ

Tytuł artykułu naukowego: *„Artificial miRNAs targeting CAG repeat expansion in ORFs cause rapid deadenylation and translation inhibition of mutant transcripts"*

Autorzy: A. Ciesiołka*, A. Stroynowska-Czerwińska*, P. Joachimiak, A. Ciołak, E. Kozłowska, M. Michalak, M. Dąbrowska, M. Olejniczak, K. Raczyńska, D. Zielińska, M. Woźna-Wysocka, W. Krzyżosiak, A. Fiszer

Czasopismo: Cellular and Molecular Life Sciences

Oświadczam, że mój wkład w wyżej wymieniony artykuł naukowy polegał na:

- Udziale w wygenerowaniu konstruktów genetycznych zawierających sekwencje egzonu 1 *HTT* oraz dwóch genów reporterowych
- Udziale w wyprowadzeniu stabilnych linii komórkowych HEK Flp-In T-REx-293 z indukowalną ekspresją w/w konstruktów genetycznych
- Przeprowadzeniu transfekcji, pomiarów chemiluminescencji oraz analizy qPCR w celu poznania kinetyki deregulacji transkryptu i białka w odpowiedzi na działanie oligonukleotydów wyciszających
- Przeprowadzeniu analiz RT-qPCR oraz G/I *tailing* do zbadania procesu deadenylacji po transfekcji oligonukleotydem wyciszającym
- Udziale w poznaniu roli białka AGO2 w mechanizmie wyciszania ekspresji badanego genu poprzez wykonanie eksperymentów transfekcji, *western blot* oraz RT-qPCR
- Opracowaniu części wyników i przygotowaniu niektórych rycin do publikacji

Joachimiak

Podpis kandydata

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, 13.03.2023r.

Paweł Joachimiak
Zakład Biotechnologii Medycznej
Instytut Chemii Bioorganicznej PAN
ul. Noskowskiego 12/14

OŚWIADCZENIE KANDYDATA O WŁASNYM WKŁADZIE W PUBLIKACJE NAUKOWE WCHODZĄCE W SKŁAD ROZPRAWY DOKTORSKIEJ

Tytuł artykułu naukowego: *„ Implications of Poly(A) Tail Processing in Repeat Expansion Diseases"*

Autorzy: P. Joachimiak, A. Ciesiołka, G. Figura, A. Fiszer

Czasopismo: Cells

Oświadczam, że mój wkład w wyżej wymieniony artykuł naukowy polegał na:

- Zebraniu i analizie dostępnej literatury na temat poliadenylacji oraz alternatywnej poliadenylacji, metod wykorzystywanych do określania długości ogonów poli(A) oraz predykcji miejsc alternatywnej poliadenylacji, powiązań alternatywnej poliadenylacji z chorobami spowodowanymi ekspansją powtórzeń tandemowych
- Przeanalizowaniu dostępnych baz danych zawierających informacje odnośnie miejsc poliadenylacji i wiązania miRNA w transkryptach chorób poliglutaminowych
- Wykonaniu wszystkich rycin i tabel do publikacji oraz przygotowaniu materiałów suplementarnych
- Pisaniu tekstu publikacji oraz przygotowaniu odpowiedzi i poprawek na otrzymane recenzje

Joachimiak

Podpis kandydata

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, 13.03.2023r.

Paweł Joachimiak
Zakład Biotechnologii Medycznej
Instytut Chemii Bioorganicznej PAN
ul. Noskowskiego 12/14

OŚWIADCZENIE KANDYDATA O WŁASNYM WKŁADZIE W PUBLIKACJE NAUKOWE WCHODZĄCE W SKŁAD ROZPRAWY DOKTORSKIEJ

Tytuł artykułu naukowego: *„Allele-specific quantitation of ATXN3 and HTT transcripts in polyQ disease models"*

Autorzy: P. Joachimiak, A. Ciesiołka, E. Kozłowska, P. M. Świtoński, G. Figura, A. Ciołak, G. Adamek, M. Surdyka, Ż. Kalinowska-Pośka, M. Figiel, N. S. Caron, M. R. Hayden, A. Fiszer

Czasopismo: BMC Biology

Oświadczam, że mój wkład w wyżej wymieniony artykuł naukowy polegał na:

- Udziale w zidentyfikowaniu wariantów SNP występujących w sekwencjach kodujących badanych genów i przypisaniu ich do odpowiedniego allelu
- Wyborze metody mającej służyć do ilościowego określenia poziomu ekspresji obu alleli każdego badanego genu
- Prowadzeniu hodowli fibroblastów SCA3 i zbiorze materiału komórkowego
- Wykonaniu wszystkich eksperymentów z użyciem metody ddPCR, wraz z optymalizacją metody oraz analizą i opracowaniem wyników
- Udziale w przeprowadzeniu transfekcji komórek NSC HD oligonukleotydami wyciszającymi
- Wykonaniu wszystkich rycin do publikacji i przygotowaniu materiałów suplementarnych
- Pisaniu tekstu publikacji oraz opracowaniu odpowiedzi na recenzje

Joachimiak

.....

Podpis kandydata

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, 13.03.2023r.

Dr hab. Agnieszka Fiszer, prof. ICHB PAN
Zakład Biotechnologii Medycznej
Instytut Chemii Bioorganicznej PAN
ul. Noskowskiego 12/14, Poznań

OŚWIADCZENIE AUTORA DO KORESPONDENCJI O WKŁADZIE MGR. PAWŁA JOACHIMIAKA W PUBLIKACJE NAUKOWE WCHODZĄCE W SKŁAD JEGO ROZPRAWY DOKTORSKIEJ

„Artificial miRNAs targeting CAG repeat expansion in ORFs cause rapid deadenylation and translation inhibition of mutant transcripts"
A. Ciesiołka, A. Stroynowska-Czerwińska, P. Joachimiak, A. Ciołak, E. Kozłowska, M. Michalak, M. Dąbrowska, M. Olejniczak, K. Raczyńska, D. Zielińska, M. Woźna-Wysocka, W. Krzyżosiak, A. Fiszer
Cellular and Molecular Life Sciences

Wkład Doktoranta polegał na udziale w: otrzymaniu konstruktów genetycznych, wyprowadzeniu linii komórkowych (kluczowych do uzyskania głównych wniosków w niniejszej pracy), badaniu kinetyki deregulacji transkryptu i białka w odpowiedzi na działanie oligonukleotydów wyciszających, określeniu roli białka AGO2 w mechanizmie wyciszania. Ponadto ważną część badań stanowiły analizy kinetyki skracania ogonów poliA, w ramach których Doktorant przeprowadził analizy eksperymentalne oraz opracował wyniki do publikacji.

„Implications of Poly(A) Tail Processing in Repeat Expansion Diseases"
P. Joachimiak, A. Ciesiołka, G. Figura, A. Fiszer
Cells

Wkład Doktoranta jako pierwszego autora tej publikacji polegał na: udziale w opracowaniu koncepcji publikacji, przeglądzie dotychczasowej literatury, wykonaniu analiz danych dotyczących alternatywnych miejsc poliadenylacji w transkryptach związanych z chorobami poliglutaminowymi, opracowaniu wszystkich rycin, przygotowaniu pierwszej wersji manuskryptu oraz pracy nad finalną jego wersją i odpowiedziami na recenzje.

„Allele-specific quantitation of ATXN3 and HTT transcripts in polyQ disease models"
P. Joachimiak, A. Ciesiołka, E. Kozłowska, P. M. Świtoński, G. Figura, A. Ciołak, G. Adamek, M. Surdyka, Ż. Kalinowska-Pośka, M. Figiel, N. S. Caron, M. R. Hayden, A. Fiszer
BMC Biology

Wkład Doktoranta jako pierwszego autora tej publikacji polegał na: udziale w identyfikacji wariantów SNP, wyborze metody i zestawów do badań, udziale w otrzymywaniu materiału z linii komórkowych, przygotowaniu matryc i wykonaniu wszystkich reakcji ddPCR, analizie wyników, przygotowaniu rycin do publikacji, jak i pierwszej wersji manuskryptu oraz pracy nad finalną jego wersją i odpowiedziami na recenzje.

Agnieszka Fiszer
.....

Podpis

Pełne teksty publikacji wchodzących
w skład rozprawy doktorskiej

Cellular and Molecular Life Sciences
https://doi.org/10.1007/s00018-020-03596-7

ORIGINAL ARTICLE

Artificial miRNAs targeting CAG repeat expansion in ORFs cause rapid deadenylation and translation inhibition of mutant transcripts

Adam Ciesiolka[1] · Anna Stroynowska-Czerwinska[1,2] · Paweł Joachimiak[1] · Agata Ciolak[1] · Emilia Kozlowska[1] · Michal Michalak[1] · Magdalena Dabrowska[3] · Marta Olejniczak[3] · Katarzyna D. Raczynska[4,5] · Dominika Zielinska[1] · Magdalena Wozna-Wysocka[1] · Wlodzimierz J. Krzyzosiak[1] · Agnieszka Fiszer[1]

Received: 30 December 2019 / Revised: 1 July 2020 / Accepted: 9 July 2020

Abstract

Polyglutamine (polyQ) diseases are incurable neurological disorders caused by CAG repeat expansion in the open reading frames (ORFs) of specific genes. This type of mutation in the *HTT* gene is responsible for Huntington's disease (HD). CAG repeat-targeting artificial miRNAs (art-miRNAs) were shown as attractive therapeutic approach for polyQ disorders as they caused allele-selective decrease in the level of mutant proteins. Here, using polyQ disease models, we aimed to demonstrate how miRNA-based gene expression regulation is dependent on target sequence features. We show that the silencing efficiency and selectivity of art-miRNAs is influenced by the localization of the CAG repeat tract within transcript and the specific sequence context. Furthermore, we aimed to reveal the events leading to downregulation of mutant polyQ proteins and found very rapid activation of translational repression and *HTT* transcript deadenylation. Slicer-activity of AGO2 was dispensable in this process, as determined in *AGO2* knockout cells generated with CRISPR-Cas9 technology. We also showed highly allele-selective downregulation of huntingtin in human HD neural progenitors (NPs). Taken together, art-miRNA activity may serve as a model of the cooperative activity and targeting of ORF regions by endogenous miRNAs.

Keywords miRNA · CAG repeats · Polyglutamine diseases · Huntington's disease · Translational inhibition

This work is dedicated in the memory of Wlodzimierz Krzyzosiak, deceased in December 2017, who extensively developed research on repetitive sequences and his wife, Krystyna Krzyzosiak, deceased in May 2019, who greatly supported our scientific work.

Adam Ciesiolka and Anna Stroynowska-Czerwinska Joint Authors.

Wlodzimierz J. Krzyzosiak: Deceased.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00018-020-03596-7) contains supplementary material, which is available to authorized users.

✉ Agnieszka Fiszer
agnieszka.fiszer@ibch.poznan.pl

[1] Department of Molecular Biomedicine, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, Poznan, Poland

[2] Laboratory of Structural Biology, International Institute of Molecular and Cell Biology, Ks. Trojdena 4, Warszawa, Poland

[3] Department of Genome Engineering, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, Poznan, Poland

[4] Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University in Poznan, Wieniawskiego 1, Poznan, Poland

[5] Center for Advanced Technology, Adam Mickiewicz University, Wieniawskiego 1, Poznan, Poland

Abbreviations

16CAG cell line	Stably expressing exon 1 of *HTT* with 16 CAG repeats
98CAG cell line	Stably expressing exon 1 of *HTT* with 98 CAG repeats
art-miRNAs	CAG repeat-targeting artificial miRNAs
ASO	Antisense oligonucleotide
ATN1	Atrophin-1
ATXN3	Ataxin-3
ATXN7	Ataxin-7
ddPCR	Digital droplet PCR
DRPLA	Dentatorubral–pallidoluysian atrophy

Published online: 21 July 2020

Fluc	Firefly luciferase
HTT	Huntingtin
HD	Huntington's disease
ICC	Immunocytochemistry
iPSCs	Induced pluripotent stem cells
miRNA	MicroRNA
NP	Neural progenitor
ORF	Open reading frame
polyQ	Polyglutamine
RISC	RNA-induced silencing complex
SCA3	Spinocerebellar ataxia type 3
sgRNA	Small guide RNA
siRNA	Short interfering RNA
smFISH	Single-molecule fluorescent in situ hybridization
Rluc	*Renilla* Luciferase
NlucP	Nano luciferase with PEST domain
RT-qPCR	Quantitative reverse transcription PCR
UTR	Untranslated region

Introduction

Non-coding RNAs (ncRNAs) are a large, diverse group of transcripts that do not contain information about protein sequence but mainly play a crucial role in the post-transcriptional regulation of gene expression. Examples of ncRNAs are short interfering RNAs (siRNAs) and microRNAs (miRNAs), which constitute a large family of short (~21 nt) RNAs [1–4]. SiRNAs activate RNA-induced silencing complex (RISC) to carry out the AGO2-mediated cleavage of a transcript within a perfectly matched siRNA-mRNA duplex, followed by mRNA degradation [5]. In contrast, the miRNA strand guides the miRNA-induced silencing complex (miRISC) to interact with only partially complementary sequences within the transcripts in animal cells, causing translational inhibition and mRNA transcript decay following deadenylation [6–10].

Functional miRNA-binding sites are usually localized within the 3′ untranslated region (UTR) but might also be present within the open reading frame (ORF) [11–15] and 5′UTR [16–18]. These latter sites are considered as less functional than those in the 3′UTR as miRISCs cannot avoid collision with the scanning small ribosomal subunit and rapidly translocating ribosomes [19]. Importantly, the efficiency of the miRNA-mediated regulation of gene expression may depend on the number of miRNA-binding sites within the regulated target [20] and the distance between these sites [21]. The more target sites at an optimal distance on mRNA there are, the higher the observed inhibitory effect is, caused by cooperative interaction between miRISCs bound to neighboring sites [22].

SiRNAs and miRNAs, as negative regulators of gene expression, are often used in the development of therapeutic approaches. One of the examples are strategies for incurable and progressive neurodegenerative polyglutamine (polyQ) diseases which include Huntington's disease (HD), spinal bulbar muscular atrophy (SBMA), dentatorubral–pallidoluysian atrophy (DRPLA) or spinocerebellar ataxia (SCA) types 1, 2, 3, 6, 7 and 17 (Fig. 1a). These disorders are caused by the expansion of CAG repeat sequences within the ORFs of specific genes, so that the normal alleles contain 10–20 CAG repeats, whereas mutant alleles usually 40–70 CAG repeats. Due to a location of mutation within ORF, mutated gene encode protein with an expanded polyQ tract [23, 24].

One promising therapeutic approach is the elimination of mutant gene expression by directly targeting the mutation site in the transcript, i.e., the expanded CAG repeat tract [25, 26]. In a series of studies from David Corey's and our groups, the effects of particular oligonucleotides, hereafter called CAG repeat-targeting artificial miRNAs (art-miRNAs), were tested in various polyQ disease models [27–40] (Table S1). The common feature of these oligonucleotides is the presence of specific mismatches in the interaction with the targeted CAG repeat tract, making these oligonucleotides similar to miRNAs. Allele-selective downregulation of mutant polyQ proteins by art-miRNAs most probably results from preferential activation of the silencing mechanism when multiple miRISCs are present on the expanded repeat tract. The targeted transcript level was less affected than the protein level as art-miRNAs did not induce the substantial mRNA cleavage typical of siRNAs [27, 28, 30]. Moreover, a study of the mechanism of action of art-miRNAs suggested cooperative silencing by miRISCs located on the expanded repeat tract, as revealed by dose-response experiments, and the involvement of AGO2 and GW182, as shown by RNA immunoprecipitation and siRNA-based knockdown experiments [30]. Intriguingly, the activities of art-miRNAs in various polyQ disease models differed significantly, and minor differences in oligonucleotide sequence largely affected the observed activity (Table S1). Therefore, we decided to investigate the details of the activated silencing process in the context of further development of this approach and as an example of miRNA-based targeting of ORF regions.

In this study, we aimed to elucidate the key factors affecting silencing efficiency of art-miRNAs and determine the mechanism of their action. For this purpose we used cells with endogenous mutant gene expression, including human neural progenitors (NPs), as well as we constructed several dedicated cellular models. We compiled the results of testing our most effective art-miRNA, A2 [33], to highlight the

variance in its activity in different polyQ disease models. We show that allele-selectivity of art-miRNAs is determined by the localization of CAG repeat tract in ORF and strengthened by specific sequence of huntingtin (*HTT*) transcript. Moreover, we demonstrate that A2 induced rapid mRNA deadenylation and translation inhibition and AGO2 was not required in activated silencing mechanism.

Materials and methods

Cell lines

HEK 293T (American Type Culture Collection) and host Flp-In T-REx-293 cell lines (Thermo Fisher Scientific) were cultivated in Dulbecco's Modified Eagle's Medium (Sigma-Aldrich), containing 10% fetal bovine serum (Biowest), penicillin-streptomycin solution (Sigma-Aldrich), 2 mM L-glutamine (Sigma-Aldrich). Additionally, for 16CAG and 98CAG Flp-In T-REx-293 cell culture 100 µg/ml hygromycin B (Thermo Fisher Scientific) and 5 µg/ml blasticidin S (Thermo Fisher Scientific) was supplemented. Patient-derived fibroblasts (Coriell Institute, SCA3 GM06153: 17/70 CAG repeats in *ATXN3*; HD GM04281: 17/68 CAG repeats in *HTT*, DRPLA GM13716: 16/68 CAG repeats in *ATN1*, SCA7 GM03561: 8/62 CAG repeats in *ATXN7*; and control line GM05565) were grown in Eagle's Minimal Essential Medium (Sigma-Aldrich) supplemented with 10% fetal bovine serum (Sigma-Aldrich), antibiotic–antimycotic solution (Sigma-Aldrich), 2 mM GlutaMAX (Gibco) and MEM non-essential amino acids (Sigma-Aldrich).

Human neural progenitors (NPs) were derived from HD induced pluripotent stem cells (iPSC) ND42222 (19/109 CAG repeats in *HTT*) obtained from NINDS Human Genetics Resource Center (Coriell Institute). For neural induction STEMdiff SMADi Neural Induction Kit (STEMCELL Technologies) was used according to monolayer protocol, following manufacturer's instructions. Briefly, iPSC were grown in Essential 8 (Gibco) medium on Geltrex (Gibco) coated 6-well plate until 70–80% confluence was reached. Then, iPSCs were dissociated to single cells by incubation with 0.5 mM EDTA in PBS for 10 min. Cells were counted using TC20 Automated Cell Counter (Bio-Rad) and resuspended at 1×10^6 cells/ml density for seeding in STEMdiff Neural Induction Medium with SMADi and 10 nM Y-27632 (all from STEMCELL Technologies). For further cultivation cells were detached using Accutase (STEMCELL Technologies) and after third passage they were grown in STEMdiff Neural Progenitors Medium (STEMCELL Technologies). Expression of *SOX1*, *SOX2*, *PAX6,* and *NES* markers was confirmed by ICC (Supplementary Figure S4A) and by RT-qPCR (Supplementary Figure S4B). All cell lines were cultured at an appropriate cell confluence at 37 °C in 5% CO_2. Cell banks were stored in liquid nitrogen.

RNA oligonucleotides

All siRNA oligonucleotides (Table S2) were synthesized by Metabion or Future Synthesis, dissolved in water to 100 µM concentration and stored at − 80 °C. To obtain 20 µM duplexes sense and antisense strands were diluted in annealing buffer, heated for 1 min in 90 °C and kept for gradual cooling at room temperature for 45 min.

Transfection

Lipofectamine 2000 (Invitrogen) was used to transfect HEK 293T, Flp-In T-REx-293 cells and fibroblasts with plasmids and oligonucleotides, accordingly to the manufacturer's protocol. 24 h prior to transfection cells were plated after estimation of cell number. To optimize and monitor transfection efficiency control fluorescent BlockIT siRNA (Invitrogen) or control plasmid encoding GFP (System Biosciences) was used. Cells were harvested at specific time points indicated in figure legends. Briefly, HEK 293T line (120,000 cells/well seeded into 24-well plate) was co-transfected with 100 ng of plasmid of pmirGLO construct and 50 nM oligonucleotide using 1.5 µl Lipofectamine 2000 in 300 µl medium. Generated Flp-In T-REx-293 lines (160,000 cells/well seeded into 12-well plate) were transfected with 100 nM oligonucleotide using 4 µl Lipofectamine 2000 in 1.2 ml of medium. Transfection of NPs was performed at fourth or fifth passage using 2 µl of siPORT Amine (Ambion) per well of 6-well plate in 1 ml of complete medium. After 3 h medium was replaced with fresh one and after next 24 h the medium was changed for the media lacking Y-27632. NPs were harvested using Accutase fixed 48 h post-transfection

Protein isolation and western blot

Cells were collected at specific time points for particular experiments (which are given in Figure legends), e.g. time points selected for most efficient downregulation of specific proteins in fibroblasts and NPs. Cell pellets were washed once with PBS and lysed with PB buffer (60 mM Tris-base, 2% SDS, 10% sucrose, 2 mM PMSF). Next, the cell extract was heated in 95 °C for 5 min and protein concentration was estimated based on measurement at 280 nm using DeNovix

spectrophotometer. Equal amounts (~30 µg) of total protein were diluted in loading buffer and heated in 95 °C for 5 min and run on SDS-polyacrylamide gels: 5% stacking, 10% resolving gel in Tris/glycine/SDS buffer for ataxin-3 and luciferase detection; 3–8% NuPAGE Tris acetate gels (Thermo Fisher Scientific) in XT Tricine buffer (Bio-Rad) with cooling in ice-water bath for atrophin-1 and huntingtin detection. Next, proteins were wet-transferred to nitrocellulose membrane (GE Healthcare) and specific primary (anti-ataxin-3, anti-huntingtin, anti-atrophin-1, anti-vinculin and anti-Fluc) and horseradish peroxidase-conjugated secondary antibodies (anti-rabbit or anti-mouse) were used. All antibodies used are given in Table S3**.** The immunodetection was performed using WesternBright Quantum HRP Substrate (Advansta). The chemiluminescent signals were scanned from membranes using GBOX documentation system (Syngene) and the bands were quantified using Gel-Pro Analyzer.

RNA isolation, RT-qPCR and ddRT-PCR

After cell lysis in TRI Reagent (ThermoFisher), Direct-zol RNA MiniPrep kit (ZymoResearch) or Total RNA Zol-Out kit (A&A Biotechnology) was used for total RNA isolation. For Flp-In T-REx-293 cell lines, a fraction of lysates prepared in Cytoplasmic Lysis Buffer [PBS, 0.1% NP40, cOmplete EDTA-free Protease Inhibitor Cocktail (Roche)] was mixed with four volumes of TRI Reagent for further isolation. The concentration of isolated total RNA was assessed by measurement at 260 nm using DeNovix spectrophotometer. Reverse transcription was performed using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) and random hexamer primers (Promega), according to the manufacturer's protocols. RT-qPCR was performed using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) and CFX Connect Real-Time System (Bio-Rad), according to the manufacturer's protocols and established guidelines for qPCR. Digital droplet PCRs (ddPCRs) were prepared using DG8 cartridges and gaskets, QX200 Droplet Generation Oil and QX200 EvaGreen Digital PCR Supermix (BioRad) and performed on QX200 Droplet Digital PCR System (BioRad), according to the manufacturer's protocols. All primer sequences are listed in Table S4.

Hill coefficient calculation

The obtained results of an average relative protein level have been fitted by GraphPad Prism to the Hill equation curve ($y=a+(b-a)/[1+(K/x)^N]$, where x is oligonucleotide concentration, y is relative protein expression, a is minimal value of y, b is maximal value of y, K is fitting parameters and N determines the slope of the curve and the value is the Hill coefficient, nH).

smFISH

Probes, buffers and protocol from Stellaris RNA FISH technology (Biosearch Technologies) were used. Probes 3′-labelled with Quasar 670 dye were used for human *HTT* (cat # SMF-20836-5) and *ATXN3* (Custom Stellaris RNA FISH probes designed using online Stellaris probe designer, sequences are listed in Table S5), and with CAL Fluor 590 dye for *GAPDH* (cat # SMF-2026-1). Cells were fixed in 4% paraformaldehyde in PBS for 20 min at RT, then prehybridized in Wash Buffer A containing 10% formamide for 5 min at RT. Hybridization was performed in Hybridization Buffer with 10% formamide at 37 °C overnight. Washing was performed with Wash Buffer A for 30 min at 37 °C and next with Wash Buffer B for 5 min at RT. SlowFade Diamond Antifade Mountant (Thermo Fisher Scientific) was used for nuclear staining. Images were captured with Leica DMI6000 inverted fluorescence microscope equipped with DFC360 FX camera. Excitation/emission filters sets were Leica A for DAPI, Chroma 49005 and 49009 for CAL Fluor 590 and Quasar670, respectively. To visually examine data, a maximum intensity z-projection of all of slices in each stack were created using ImageJ. Signals were simplistically attributed as nuclear based on DAPI staining. Quantification of individual RNA FISH spots was done using the StarSearch software (https://www.seas.upenn.edu/~rajlab/StarSearch/launch.html).

Luciferase-based plasmids containing CAG repeat tracts

The plasmids were generated on the basis of the pmirGLO Vector (Promega) encoding Firefly luciferase (Fluc) and Renilla luciferase (Rluc). CAG repeat tract sequences were inserted into *Fluc* gene, either downstream ("3′UTR" and "3′ORF" plasmids; between SalI and XbaI restriction sites) or upstream ("5′ORF" plasmids; between restriction sites for EcoRI and NdeI inserted in two steps using QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies)). "3′ORF" plasmid was generated by mutation of "3′UTR" plasmid in the *Fluc* gene STOP codon. *ATXN3* and *HTT*-specific inserts with normal and mutant CAG repeat tracts were obtained by PCR using cDNA from fibroblast cell lines (SCA3 and HD) and primers with specific restriction sites. Short synthetic inserts (containing 17 CAG repeats and including specific restriction sites) were chemically synthesized (Sigma-Aldrich) and annealed for cloning. For longer synthetic inserts we used in vitro repeat expansion method known as Synthesis of Long Iterative Polynucleotide (SLIP) [41, 42] with the "17CAG" insert as initial template for the repeat tract expansion. After obtaining first plasmid

with expanded CAG tract, we further modified a protocol and used two plasmids with various lengths of CAG tract in SLIP. This approach allowed for generating longer expansion at one step. Due to DNA polymerase slipping the lengths of mutated constructs are slightly different. Ligation of inserts with plasmids was performed using T4 Ligase (Promega) according to the manufacturer's procedure. Next, competent DH5α *E. coli* cells were transformed and plasmids were isolated using Endotoxin-free MidiPrep kit (Qiagen). Due to technical problems mutant synthetic inserts are not included for the set of "5′ORF" constructs. Sequences of DNA oligonucleotides used for cloning are given in Table S6.

Flp-In T-REx-293 cell lines for inducible and stable expression of CAG repeat tracts

All components of the designed dual-luciferase system were cloned into pcDNA5/FRT/TO vector (Invitrogen) for obtaining inducible expression in Flp-In T-REx-293 cell lines of either normal or mutant *HTT* fragment. This vector was integrated into the genome via Flp recombinase-mediated DNA recombination at the FRT site. Sequences of *Fluc* and Nano luciferase with PEST domain (*NlucP*) were cloned from pmirGLO and pNL1.2 (Promega) vectors, respectively. A bidirectional inducible promoter system (BI-16) capable of reproducible coexpression of two proteins was constructed [43]. Additional SV40pA sequences were cloned from pNL1.2 (Promega) at respective sites. The full exon 1 of *HTT* containing either 16 or 98 CAG repeats was amplified in PCR and these inserts were cloned upstream of *NlucP* sequence to obtain expression of *HTT-NlucP* fusion gene. Sequences of DNA oligonucleotides used for cloning are given in Table S7. The pcDNA5/FRT/TO-based expression constructs and pOG44 vector were co-transfected (at 1:9 ratio) into Flp-In T-REx-293 host cells using Lipofectamine 2000, according to manufacturer's protocol. Selection of hygromycin-resistant monoclones that contain stably integrated expression cassette was performed using 100 µg/ml hygromycin B, according to manufacturer's protocols. Additional details on the DNA cloning procedure and the generation of these stable cell lines are given in Supplementary Methods.

Luciferase assay

For the results presented in the Figs. 2 and S2 assays were performed using Dual-Luciferase Reporter Assay System (Promega), accordingly to the manufacturer's protocol. Briefly, cells were lysed 24 h after transfection in Passive Lysis Buffer (Promega), followed by the luciferase activity measurement using Centro LB 960 Luminometer (Berthold Technologies). The Fluc measurement data was normalized firstly to Rluc signal in the sample, next to Fluc/Rluc signal ratio obtained in cells co-transfected with negative control (NTC, non-targeting siRNA) and particular plasmid, and finally to Fluc/Rluc signal ratio measured for cells treated with plasmid lacking CAG repeat insert and particular oligonucleotide.

For the results presented in the Figs. 4 and 5 cells were lysed using Cytoplasmic Lysis Buffer [PBS, 0.1% NP40, cOmplete EDTA-free Protease Inhibitor Cocktail (Roche)] according to the REAP method [44]. Next, the lysates was used in the Nano-Glo Dual-Luciferase Reporter Assay System (Promega) and measured with Victor X4 Multilabel Plate Reader (Perkin Elmer), according to the manufacturer's instructions. Background values of NlucP and Fluc signals at $t = 0$ h were subtracted. Nluc signals for A2- or siHTT-treated cells were normalized to NlucP measurement obtained at specific time points for siRluc-treated cells. Finally, NlucP signal was normalized to Fluc signal in a respective sample.

Polysome profiling

The protocol was adapted from [45, 46]. Briefly, 5×10^6 Flp-In T-REx-293 16 CAG or 98CAG cells were seeded into 55 cm^2 plate in medium without antibiotics. After 24 h cells were transfected using selected oligonucleotides at final concentration of 100 nM. After additional 12 h *HTT* exogene expression was induced using 1 µg/ml doxycycline (Sigma-Aldrich) and after 3 h 100 µg/ml cycloheximide (Sigma-Aldrich) was added to inhibit translation elongation and fix ribosomes on transcripts. After 5 min of incubation at 37 °C cells were washed with ice-cold PBS containing cycloheximide and harvested in 1.5 ml of this buffer by scraping. Cells were collected by centrifugation at 300 rpm for 5 min at 4 °C and lysed in 500 µl ice-cold lysis buffer (10 mM HEPES pH 7,9; 1.5 mM $MgCl_2$; 10 mM KCl; 0.5 mM DTT; 1% Triton X-100, 100 µg/ml cycloheximide) containing also 100 µ/ml of RNasin (Promega). After 10 min incubation on ice lysates were centrifuged at 1500*g* for 5 min at 4 °C. Supernatant was collected and OD was measured at 260 nm.

10–60% sucrose gradients were prepared using Gradient Station (BioComp) in buffer containing 100 mM KCl, 20 mM HEPES pH 7.6; 5 mM $MgCl_2$, 100 µg/ml cycloheximide; 5 µ/ml RNasin and Protease Inhibitor Cocktail (Roche). 10 OD was loaded onto cooled sucrose gradients and centrifuged at 39,000 rpm for 2 h and 40 min at 4 °C using ultracentrifuge and SW 41Ti rotor (Beckman Coulter) About twenty 0.5 ml fractions were collected using Piston Gradient Fractionator (BioComp). Next, 0.5 ml of TRI-reagent (Ambion) was added to each fraction and subsequently RNA was isolated, including treatment with DNase I. Equal volumes of total RNA were reverse transcribed and *HTT-NlucP* and *GAPDH* expression levels were determined by qRT-PCR.

Poly(A) tail length measurements

The analysis was performed based on a polyG/I extension method [47] using the Poly(A) Tail-Length Assay Kit (Thermo Fisher Scientific). In these experiments 5 µg/ml actinomycin-D (Sigma-Aldrich) was added to the medium of 16CAG and 98CAG Flp-In T-REx-293 cells to stop transcription. 200 ng of isolated RNA from selected time points were taken for poly(A) tail length analysis that was performed following manufacturer's protocol. Specific primers used are listed in Table S4. For estimation of poly(A) tail lengths, a product obtained using gene-specific reverse primer was used as a reference. PCR products were analyzed on 2100 Bioanalyzer using DNA 1000 Kit (Agilent).

Generation of AGO2 knockout and AGO2(D597A) mutant stable cell lines and transient AGO2 overexpression

CRISPR-Cas9-mediated AGO2 knockout and AGO2(D597A) mutant cell lines were established using previously generated Flp-In T-REx-293 98CAG cells. For AGO2 knockout Cas9_sg1 and Cas9_sg2 plasmids, encoding sgRNA1 and sgRNA2 which are specific for target sequences within exon 2 of *AGO2* gene, were used. Cas9_sg3 encoding sgRNA3, binding to sequence within exon 14, was used for AGO2(D597A) mutant cell line. To generate these plasmids, sense and antisense DNA strands of sgRNAs were annealed and ligated into pSpCas9(BB)-2A-GFP (PX458) (Addgene) plasmid, digested with the FastDigest BpiI (Thermo Fisher Scientific). Chemically competent *E. coli* GT116 cells (InvivoGen) were transformed with the 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. For nucleofection Flp-In T-REx-293 98CAG cells were electroporated with the Neon Transfection System (Invitrogen). Briefly, 1×10^5 cells were harvested, resuspended in Buffer R and electroporated with 1 µg of plasmid DNA (500 ng of each Cas9_sg1 and Cas9_sg2 plasmids) in 10 µl tips using the following parameters: 1100 V, 20 ms, two pulses. For AGO2(D597A) cell line generation, cells were electroporated with 1 µg of plasmid DNA and 1 µl of 100 µM single-stranded donor oligonucleotide (ssODN) (IDT) harboring GAC to GCC codon change. Selection of clones is described in Supplementary Materials and Methods. The oligonucleotide sequences are included in Table S8. For AGO2 overexpression, AGO2 coding sequence was amplified using PCR from pIRESneo FLAG/HA Ago2 plasmid (Addgene, #10822) [48] and cloned into pcDNA3.1(+) (Invitrogen) between HindIII/BamHI sites.

Statistical analysis

Analyses were performed using GraphPad Prism software. Two-tailed p value < 0.05 was considered significant and is depicted on the graphs by: $*0.05 > p > 0.01$; $**0.01 > p > 0.001$; $***p < 0.001$. All experiments which resulted in statistically-analyzed quantitative data were repeated at least three times (the exact number of biological replicates, n, is given in figure legends). Depending on the experimental setup specific statistical tests were used and are indicated in figure legends. The error bars in the graphs represent standard deviations.

Results

A2-mediated silencing of different polyQ disease-related genes is varied

A large set of art-miRNAs have been tested in fibroblasts derived from patients with several polyQ diseases (Table S1). We have now complemented our results obtained with the A2 oligonucleotide [33, 38, 39] and present a direct comparison of its activity when used at the same concentration in cell lines bearing similar repeat tract lengths in mutant alleles, i.e., 62–70 CAG repeats (Fig. 1a). Overall, we observed that the efficiency and allele-selectivity of A2 art-miRNA differed in various models of polyQ diseases. The highest degree of allele-selectivity was achieved for the downregulation of *HTT* and *ATXN7*, where mutant huntingtin and ataxin-7 proteins were lowered to ~20% of control level, without reduction in normal protein levels. For ataxin-7 we also observed significant increase in normal protein level after A2 treatment [39]. Normal *ATN1* and *ATXN3* alleles were more susceptible to downregulation, but levels of normal atrophin-1 and ataxin-3 was decreased by no more than 50% of the control level with a relatively high concentration of A2 (50 nM) (Fig. 1a). In all of the examined disease models, A2 caused an allele-selective decrease in mutant protein levels at a wider range of concentrations used, as shown in the DRPLA model example (Fig. 1b). Mutant atrophin-1 was downregulated with 20 nM A2 to less than 20% of the control level without a reduction in normal protein level. Additionally, we analyzed the activity of A2 at a very wide range of concentrations in HD and SCA3 models to assess potential cooperative activity, as was previously reported for other art-miRNAs [30, 31]. We obtained a Hill coefficient (nH) value considerably > 1 (~1.7) what suggests cooperative activity of the silencing machinery for the downregulation of mutant huntingtin (Figs. 1c, S1A). On the other hand, among the models investigated, the ataxin-3 protein was decreased in the least allele-selective manner, and the obtained nHs suggest cooperative silencing of both

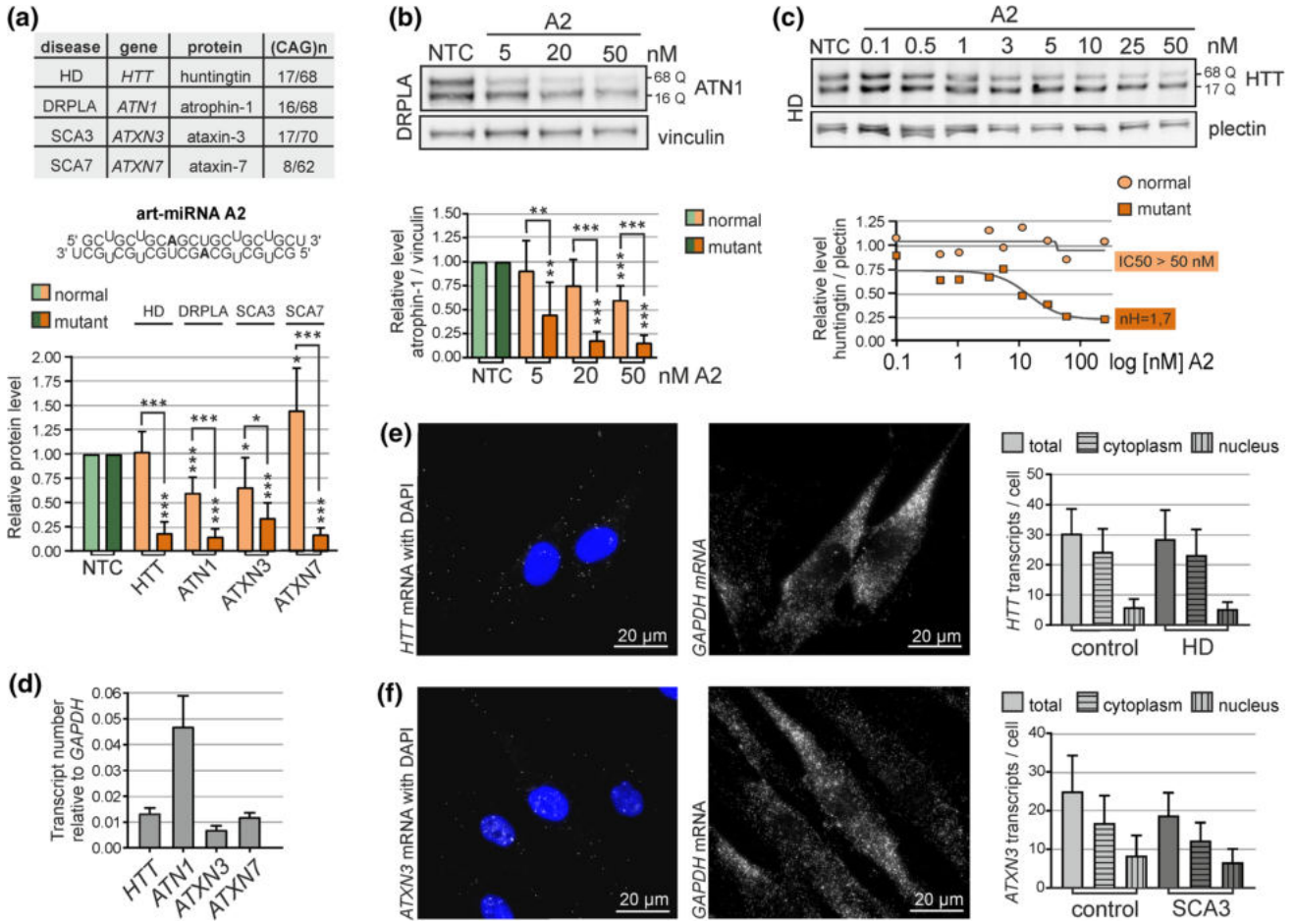
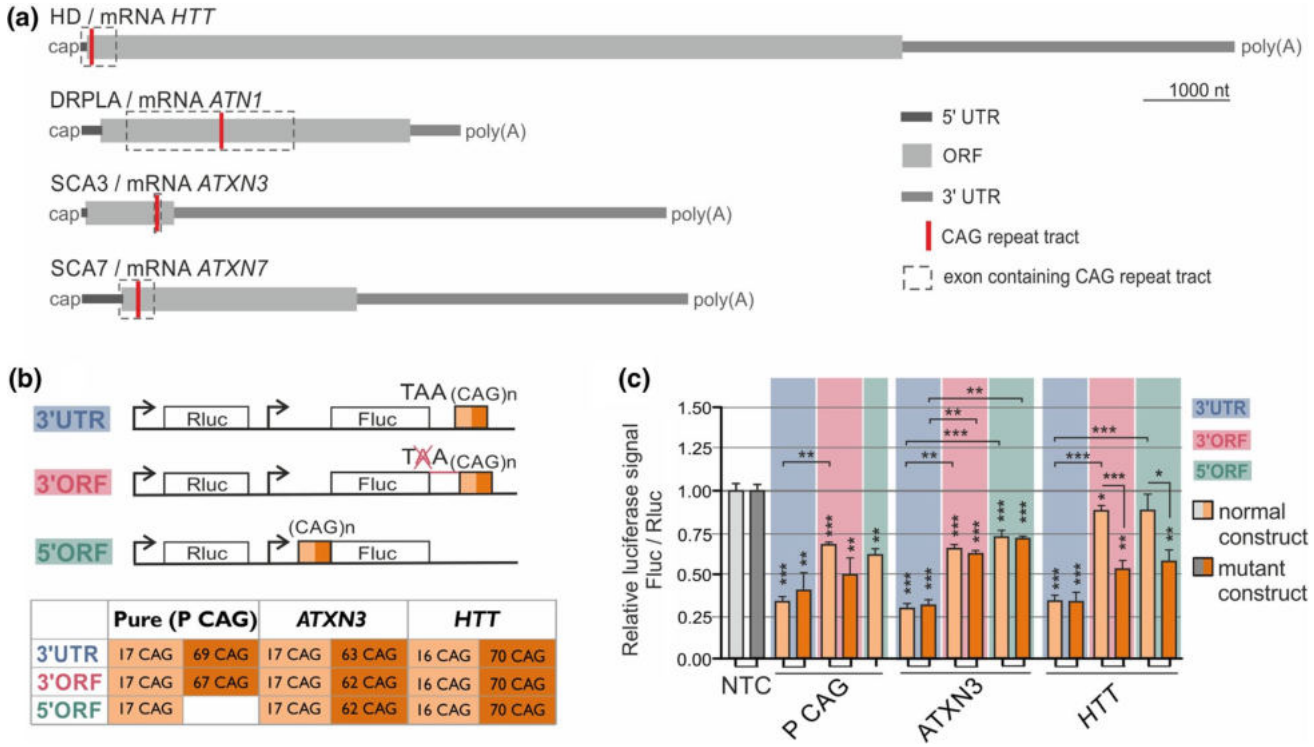


Fig. 1 A2 activity in patient-derived fibroblasts and the characteristics of targeted transcripts. **a** Upper panel: table with information about investigated models of polyQ diseases. In the last column CAG repeat tract lengths (normal/mutant allele), present in fibroblasts cells used, are given. Middle panel: sequence of art-miRNA A2 and predicted base-pairing of two strands within a duplex. Lower panel: results compiled from the western blot analysis showing HTT, ATN1, ATXN3 or ATXN7 protein levels in HD, DRPLA, SCA3 or SCA7-patient-derived fibroblasts, respectively, after transfection with 50 nM A2. Vinculin, GAPDH and plectin were used as reference proteins. NTC—cells treated with non-targeting siRNA. The results of HTT and ATXN7 downregulation are from published studies [33, 39]. The following statistical tests were used: one-sample *t* test with a hypothetical value = 1 for allele expression level; unpaired *t* test with Welch's correction for comparisons of normal and mutant allele expression. $n = 3$ **b** Western blot analysis of atrophin-1 levels in DRPLA-patient-derived fibroblasts lysed 48 h after transfection with 5, 20 or 50 nM A2. NTC—cells treated with non-targeting siRNA. Data were analyzed using one-way ANOVA (Bonferroni multiple comparisons test). $n = 3$. **c** Western blot analysis of huntingtin levels in HD fibroblasts lysed 72 h after transfection with the indicated concentration of A2. The results are presented as dose–response curves that were used to calculate the indicated Hill coefficient. NTC—cells treated with non-targeting siRNA. See Figure S1A for more data. $n = 3$ **d** Non-allele-specific quantification of *HTT*, *ATN1*, *ATXN3* and *ATXN7* transcripts with ddPCR. The results were obtained from two sets of cDNA from independent cultures of each of five fibroblast cell lines. See Figure S1C for more data. **e**, **f** Representative smFISH images for *HTT* (E) and *ATXN3* (**f**) mRNAs in HD patient- and SCA3 patient-derived fibroblasts, respectively. DAPI was used for nuclear staining. Middle panels: *GAPDH* transcripts detected in the same cells. Right panels: non-allele-specific quantification of *HTT* and *ATXN3* signals in healthy and patient-derived fibroblasts. Signals were counted from at least 50 cells. NTC used in experiments presented in this figure was BlockIT siRNA

the normal and mutant alleles by A2 (Fig. S1B). Overall, these and previous observations clearly show the targeted transcript-dependent activity of art-miRNAs.

The cellular level of polyQ disease-related transcripts is relatively low

To characterize several polyQ disease-related transcripts in more detail, we first performed their quantification in patient-derived fibroblast cells using ddPCR. The respective transcripts were present at very low levels relative to



	Pure (P CAG)		ATXN3		HTT	
3'UTR	17 CAG	69 CAG	17 CAG	63 CAG	16 CAG	70 CAG
3'ORF	17 CAG	67 CAG	17 CAG	62 CAG	16 CAG	70 CAG
5'ORF	17 CAG		17 CAG	62 CAG	16 CAG	70 CAG

Fig. 2 Impact of repeat tract length, location and sequence surrounding the targeted region on efficiency of downregulation by A2. **a** Scheme of *HTT* (NM_002111.8), *ATN1* (NM_001007026.2), *ATXN3* (NM_004993.6) and *ATXN*7 (NM_000333.3) transcripts with CAG repeat tract locations, UTRs and ORFs marked. **b** Scheme of dual-luciferase-based constructs containing the *Fluc* gene fused with the indicated CAG repeat tracts in various locations: the 3′UTR and 3′ and 5′ sites of the ORF. The exact lengths of the tracts in constructs containing CAG repeats without ('Pure') or with gene-specific surrounding sequences (from *ATXN3* or *HTT*) are given in a table. **c** Luciferase assay performed 24 h after cotransfection of HEK 293T cells with 50 nM A2 and 100 ng of the indicated plasmids. NTC—light gray: cells treated with pmirGLO plasmid and non-targeting siRNA, dark gray: "5′ORF"-modified pmirGLO plasmid treated with non-targeting siRNA (signal normalization details are given in Materials and Methods). Data were analyzed using one-way ANOVA (with Bonferroni multiple comparisons test). $n = 3$

GAPDH, and similar expression levels of *HTT* and *ATXN7*, slightly lower *ATXN3* levels and considerably higher *ATN1* levels were observed (Fig. 1d). Five separately analyzed fibroblast cell lines showed some variation in transcript levels but without outstanding tendency for fibroblasts with the mutation in specific gene, e.g., the *HTT* mRNA level in HD fibroblasts (Fig. S1C).

In addition, we performed single-molecule fluorescent in situ hybridization (smFISH) for the precise quantification and visualization of selected transcripts in fibroblasts. Microscopic analyses showed approximately 30 *HTT* transcripts and approximately 20 *ATXN3* transcripts per cell (Figs. 1e, f, S1D). Interestingly, the ratio of transcripts in the cytoplasm to those in the nucleus was approximately 4:1 and 2:1 for *HTT* and *ATXN3*, respectively. This suggests that larger cytoplasmic fraction of *HTT* transcripts, in comparison to *ATXN3* transcripts, is available for the activation of RISC-mediated processes in fibroblast cells. No substantial differences in specific mRNA copy number or localization were observed in the healthy cell line in comparison to mutant cell line (Figs. 1e, f, S1D). Therefore, we conclude that there are no significant differences in number or localization of normal vs. mutant variants of specific mRNA in human fibroblasts, although we were able to use only non-allele-specific quantification of transcripts by ddPCR and smFISH.

The presence of the targeted region in an ORF and an *HTT*-specific flanking sequences improve the allele-selectivity of art-miRNAs

The results in human fibroblasts demonstrated that activity of art-miRNAs is dependent on specific features of targeted transcripts (Fig. 1a). PolyQ disease-related transcripts differ considerably in the lengths of their ORFs and UTRs as well as the localization of CAG repeat tracts. In details, CAG repeat tract is located in *HTT* and *ATXN7* at the 5′ end of the ORF, while in *ATXN3* at the 3′ end of the ORF, and in *ATN1* in the middle of the ORF (Fig. 2a). Therefore, we decided to elucidate how CAG repeat tract localization and sequences

flanking CAG repeats influence silencing efficiency and allele preference by art-miRNAs. For this purpose, we generated pmirGLO-based plasmids encoding the *Fluc* gene, fused to a normal (~17 CAG repeats) or mutant (~65 CAG repeats) tract, and *Rluc* as internal reference (Fig. 2b). The repeat tract was placed at two sides of the *Fluc* ORF: at the *HTT*-like 5′ side ("5′ORF") or at the *ATXN3*-like 3′ side ("3′ORF"), as well as in the 3′UTR, which is a typical region for miRNA-binding sites. Inserts contained either pure CAG repeats (P CAG) or ~50 nt-long *HTT* or *ATXN3* mRNA sequences flanking both sites of the CAG repeat tract, giving a total of 17 constructs (Fig. 2b).

First, we confirmed the expression of fusion proteins in HEK 293T cells (Fig. S2A). Next, we co-transfected the designed plasmids with selected oligonucleotides (art-miRNAs, siRNA targeting *Fluc*-siFluc or non-targeting siRNA-NTC) and performed a dual-luciferase assay. Typical siRNA, siFluc, caused efficient reduction of expression of all the constructs, to ~10% of the control level, regardless the location of the target site (Fig. S2C). In contrast, for A2 we observed varied activity for particular constructs (Fig. 2c). The most prominent downregulation of *Fluc* by A2 was obtained for the targeted sequence location in the "3′UTR" constructs (Fig. 2c, bars with blue background). The luciferase signal was decreased to ~35% of the control level regardless of repeat tract length and the sequence flanking CAG repeats. Downregulation of constructs expression with the target sequence localized in the ORF of *Fluc* was less efficient (Fig. 2c, bars with a red and green background), however for constructs containing *HTT* flanking sequence (both "3′ORF" and "5′ORF") we observed significant allele-selectivity of A2 activity. In these cases, expression of the mutant construct was decreased to ~50% of the control level, whereas normal construct expression remained unchanged or decreased to only ~90% of the control level (Fig. 2c). Similar results were obtained for the other art-miRNAs analyzed: A4, G2 and G4 (Fig. S2b, c). Together, these observations demonstrate that the allele-selectivity of art-miRNAs was achieved only for constructs with *HTT*-specific sequences. In agreement with the results obtained in patient-derived fibroblasts (Fig. 1a), we conclude that better art-miRNAs allele-selectivity of *HTT* downregulation in comparison to *ATXN3*, is a combination of two effects: (1) increased downregulation of the mutant *HTT* allele as compared to mutant *ATXN3* and (2) decreased silencing of the normal *HTT* allele in comparison with normal *ATXN3*.

We also considered additional features of *HTT* and *ATXN3* mRNAs that could affect the discrepancy in art-miRNA allele-selectivity in HD and SCA3 models (Supplementary Text). For example, the presence of rare codons, upstream to miRNA-binding site, was shown to improve the efficiency of miRNA silencing for targets present in ORFs, possibly due to the decreased rate of translation [49]. Therefore, we analyzed the codon usage values in *HTT* and *ATXN3* transcript sequences upstream of the CAG repeat tracts (Fig. S3), but no significant differences were found for these mRNAs in this aspect (Supplementary Text).

A2 and siHTT caused a decrease in *HTT* mRNA in the cytoplasm of HD NPs

To investigate A2 activity in a more disease-relevant cell type, we included an additional model of human NP cells, derived from iPSCs. As A2 acted with high allele-selectivity for *HTT* silencing, we generated HD NPs. First, we characterized this cell line for the expression of neural stem cells markers (Fig. S4a, b) and optimized oligonucleotide delivery (Fig. S4c). Next, we investigated the efficiency of *HTT* expression silencing and changes in transcript abundance in HD NPs after transfection with *HTT*-specific siRNA (siHTT) or A2 art-miRNA. Similarly to results in HD fibroblasts (Fig. 1a), in HD NPs A2 allele-selectively downregulated mutant proteins to ~30% of the control level, without reduction in normal protein level, whereas siHTT decreased both alleles of the huntingtin protein to ~25% of the control level (Fig. 3a). Next, we performed a microscopic observation of endogenous *HTT* transcripts targeted with A2. Using non-allele-selective smFISH, we observed a cytoplasm-specific decrease in the huntingtin transcript number by ~45% after treatment with A2 and a more substantial reduction by ~70% after treatment with siHTT (Fig. 3b, c). This observation stays in agreement with the classical model of RISC activity in cytoplasm. Cellular localization of targeted transcript could affect the efficiency of its targeting by miRNA or siRNAs. In recent study, a larger fraction of *HTT* transcripts was detected in the nuclei of healthy human neuronal cells compared to non-neuronal cells [50]. In agreement with this observation, we also observed an increased ratio of nuclear to cytoplasmic *HTT* mRNAs in NPs, relative to fibroblasts (Figs. 1e, 3c). It is worth to notice that, after treatment with A2, we did not observe the retention of transcripts in the nucleus or cytoplasmic aggregation of mRNAs, that could result in decreased huntingtin synthesis and suggesting additional mechanisms of art-miRNA activity. Taken together, the results obtained in HD NPs show the therapeutic potential of A2, as its allele-selective activity was achieved not only in patient-derived fibroblasts but also in the neuronal cell line.

Kinetic analysis of transcript and protein downregulation shows the early events of translation inhibition induced by A2

To verify crucial factors affecting art-miRNAs activity and mechanistic details, we decided to include additional models with the exogenous expression of the targeted transcripts.

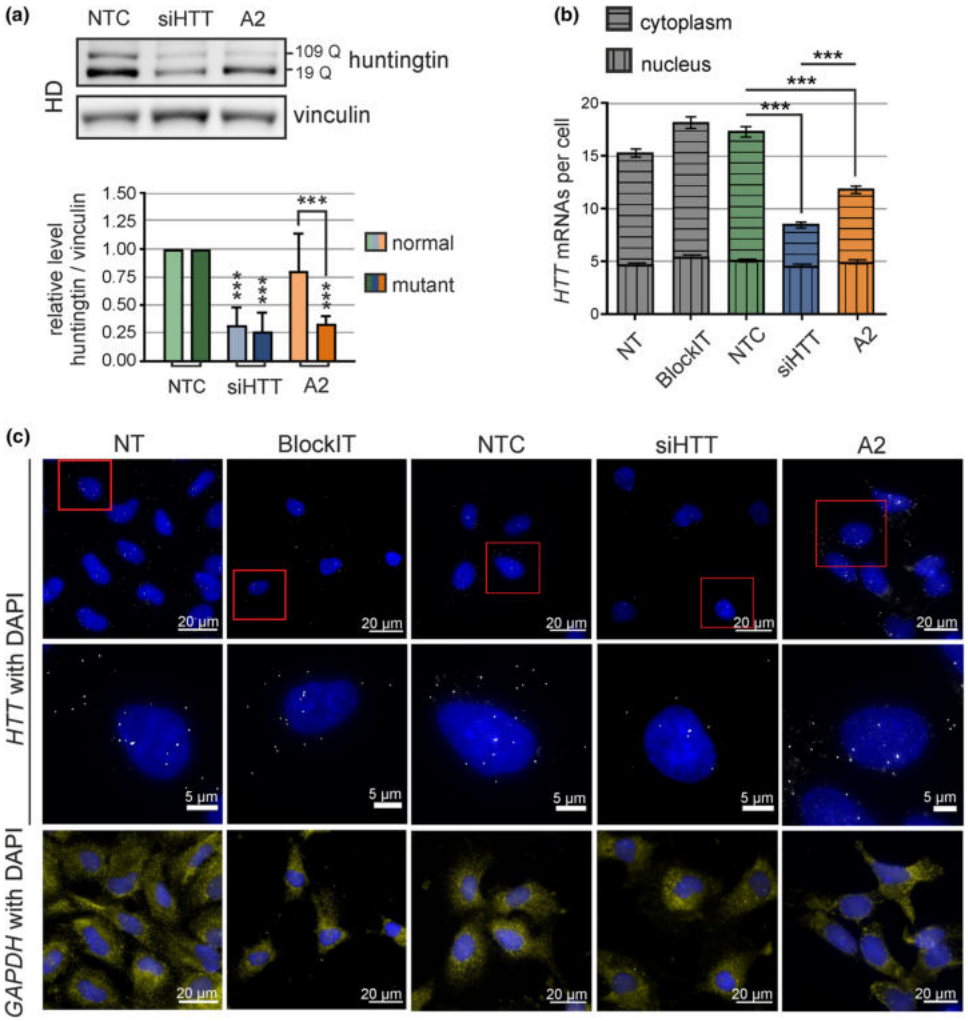


Fig. 3 A2 activity in human HD neural precursors. **a** Western blot analysis of huntingtin allele levels in HD NPs lysed 48 h after transfection with 100 nM non-targeting siRNA (NTC, BlockIT siRNA), siHTT (siRNA for *HTT*) or the art-miRNA A2. $n=4$. **b** Representative images showing the smFISH-based detection of *HTT* and *GAPDH* mRNAs in non-treated (NT) HD NP cells and HD NP cells treated with 100 nM fluorescent siRNA (BlockIT), non-targeting siRNA (NTC, siRluc), siHTT or A2. Cells were fixed 48 h after transfection with the indicated oligonucleotides. **c** Quantification of smFISH images for the experiments described in **b**. Signals were counted from at least 200 cells for each treatment. Data were analyzed using one-way ANOVA (with Bonferroni multiple comparisons test)

For this purpose, we designed a dual-luciferase system for the inducible expression of *HTT* reporters. We generated Flp-In T-REx-293 cell lines stably expressing exon 1 of *HTT* (with 16 or 98 CAG repeats, hereafter called "16CAG" and "98CAG" cell lines, respectively) fused with the *NlucP* reporter, named *HTT-NlucP* (Fig. 4a). *Fluc* expression was used as a normalization control, and both reporters were placed under a bidirectional doxycycline-inducible promoter. First, we confirmed the similar expression of the reporters and non-significant Bl-16 promoter leakage in the absence of doxycycline (Fig. S5A, B). Next, we transfected the 16CAG and 98CAG cell lines with the art-miRNA A2, siRNA specific for *HTT*, siHTT, or non-targeting siRNA, followed by induction of reporter expression and subsequent analysis of the transcript and protein levels at particular time points (Fig. 4b).

The kinetics of *HTT* reporter transcript and protein downregulation by A2 were clearly different in the 16CAG and 98CAG cell lines (Fig. 4c). For the mutant *HTT* reporter transcript, we observed a maximum of ~50% downregulation starting 3 h after induction, whereas for the normal *HTT* reporter transcript, we detected only a slight decrease (Fig. 4c, upper panel). In contrast to the transcript levels, repression of the HTT-NlucP protein by A2 was more prominent, and the mutant protein level was decreased up to ~30%, while the normal protein was decreased up to ~60% of the control level at selected time points (Fig. 4c, lower panel). Interestingly, at early time points (up to 2 h post-induction), A2 significantly lowered only the level of mutant protein, suggesting that translational repression preceded mRNA decay in the allele-selective inhibition of the mutant *HTT* allele. As a reference for the typical RNAi mechanism, we performed the same analysis with siHTT. We observed rapid transcript and protein downregulation with no apparent difference in activity towards the normal and mutant alleles (Fig. 4d), suggesting the AGO2-mediated cleavage of both transcripts and, as a result, a decrease in the protein levels.

A2 causes a shift of mutant transcripts from heavy polysome fractions

To assess translation inhibition caused by art-miRNA in more detail, we performed polysome profiling analysis of *HTT* reporter transcripts. The 98CAG and 16CAG cell lines were treated with A2 or control siRNA, and lysates were prepared 3 h after the induction of luciferase expression (Fig. 4b, e, representative UV absorbance profiles, including a profile following disruption with EDTA, are shown in Fig. S6a). In 98CAG cells A2 caused a statistically significant shift in *HTT* reporter transcript distribution in the analyzed fractions, as referred to non-targeting siRNA treatment. We observed an approximately two-fold change in mutant *HTT-NlucP* transcript abundance in selected fractions, i.e., increased cosedimentation with 40S, 60S, 80S and first light polysome fraction and decreased cosedimentation with heavier polysomes, as compared to the treatment with control siRNA (Fig. 4e, upper panels). This results suggest that A2 inhibited mutant *HTT-NlucP* translation at initiation and/or early elongation step. In these experiments, some *HTT-NlucP* transcripts remained associated with heavier polysomes after A2 treatment, probably because not all *HTT* reporter transcripts were bound by this art-miRNA. Our conclusions are supported by analogous control experiments performed in the 16CAG cell line which results did not show any significant difference in the cosedimentation of *HTT-NlucP* transcripts across the collected fractions between A2- and control siRNA-treated cells (Fig. 4e, lower panels). Moreover, as expected, no significant changes in *Fluc* expression relative to *GAPDH* expression were observed in experiments using both, 98CAG and 16CAG cell lines (Fig. S6b). Together, we conclude that observed translation inhibition occurred very rapidly and efficiently for mutant transcript as a result of A2 activity.

A2 induces rapid shortening of the targeted mRNA poly(A) tail

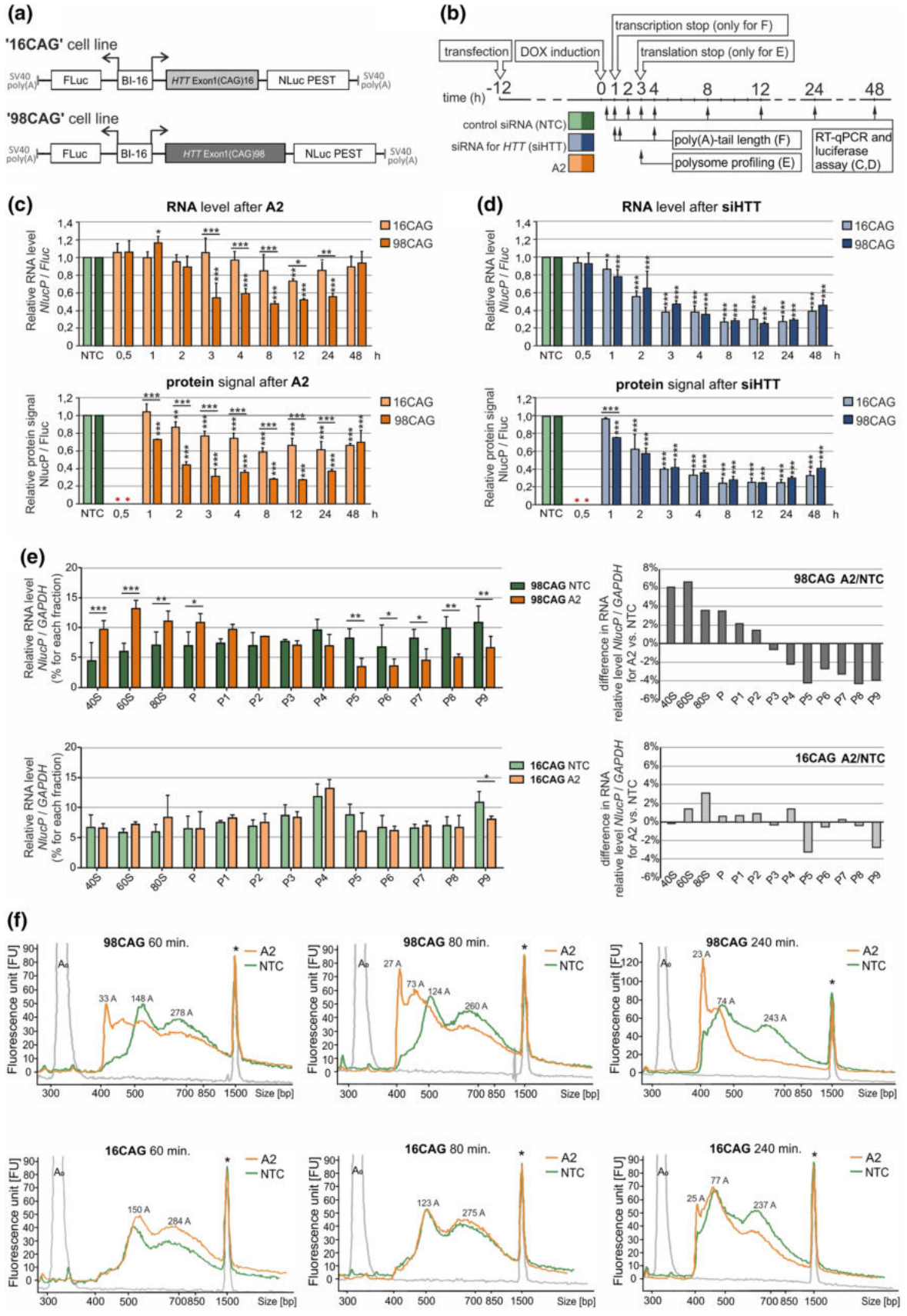
We aimed to explain in more detail the observation that mutant *HTT* reporter protein level was decreased after A2 treatment already at early time points (1–2 h) after induction, without a change in the level of its transcript (Fig. 4c). This can be explained as the effect of direct translation inhibition or transcript deadenylation, which in turn results in reduced translation due to disrupted transcript circulation. To verify the latter mechanism, we performed transcription pulse-chase experiment and examined the length of poly(A) tails in *HTT* reporter transcripts using poly G/I extension followed by resolution of the PCR products in a microfluidic chip (Fig. S7a). In details, after transfection of the 98CAG and 16CAG cell lines with A2 or control siRNA, we induced expression of the *HTT* reporter for 1 h, stopped the transcription and then analyzed poly(A) tail length profiles at three time points (Fig. 4b, f).

In the 98CAG cell line, A2 caused significant deadenylation of *HTT-NlucP* transcript already at the 60 min time point, when a substantial pool of *HTT* reporter transcript deadenylation intermediates with a short (~ 30 A) poly(A) tail appeared (Fig. 4f, upper panels). The formation of transcripts with shortened poly(A)-tails further accelerated at the 80 min and 240 min time points (Fig. 4f, upper panels). In contrast, we did not observe these deadenylation intermediates at even the 240 min time point after treatment of the 98CAG cell line with control siRNA. Analogous experiments in the 16CAG cell line showed no significant difference in the poly(A) tail length profiles of A2- and control siRNA-treated cells at the 60 and 80 min time points and only a slight difference at the latest 240 min time point (Fig. 4f, lower panels). In both cell lines and following treatment with A2 and control siRNA, we observed significant changes in the length of the poly(A) tail at 240 min compared to that at earlier time points (Fig. S7b), indicating events typical of transcript decay after transcription arrest. Taken together, A2 caused rapid deadenylation only of mutant *HTT-NlucP* transcript, suggesting crucial role of this process in the art-miRNA-mediated repression. Deadenylation is expected to lead to mutant transcript degradation observed at later time points (3–24 h) (Fig.4c).

AGO2 is dispensable for the A2-mediated silencing of *HTT* expression

Our next step was to elucidate the role of AGO2 in art-miRNA mechanism. We wanted to determine if AGO2 presence or the activity of its catalytic subunit is required for art-miRNAs-mediated silencing. To verify this, we modified the 98CAG cell line using CRISPR-Cas9 technology. We created homozygous cell lines in which *AGO2* gene was knocked out (AGO2del) or a cleavage-deficient AGO2/D597A protein was expressed (AGO2mut) (Fig. S8a, b). The D597A mutation in AGO2 is known to abolish RNA cleavage without affecting efficiency of siRNA binding or translational repression [51, 52]. Two sgRNAs were designed to knock out AGO2 by deletion of a gene fragment in exon 2 leading to premature STOP codons (Fig. S8a, b), whereas an approach using one sgRNA and a donor template containing a specific nucleotide mutation was used to introduce a catalytic mutation in AGO2 (Fig. S8c, d). We selected the final clones based on the results of DNA sequencing (Fig. S8b, d) and AGO2 immunoblotting (Fig. 5b).

To analyze the requirement of AGO2 in A2 and siHTT activities, we analyzed both the transcript and protein levels of the mutant *HTT* reporter in 98CAG-AGO2mut and -AGO-2del cell lines. In addition, we performed rescue experiments and transfected cell lines with plasmids encoding WT


(a)
'16CAG' cell line
SV40 poly(A)
FLuc
BI-16
HTT Exon1(CAG)16
NLuc PEST
SV40 poly(A)
'98CAG' cell line
SV40 poly(A)
FLuc
BI-16
HTT Exon1(CAG)98
NLuc PEST
SV40 poly(A)
(b)
transcription stop (only for F)
transfection
DOX induction
translation stop (only for E)
time (h)
-12
0 1 2 3 4
8
12
24
48
control siRNA (NTC)
siRNA for HTT (siHTT)
A2
poly(A)-tail length (F)
polysome profiling (E)
RT-qPCR and luciferase assay (C,D)
(c)
RNA level after A2
Relative RNA level NlucP / Fluc
16CAG
98CAG
NTC 0,5 1 2 3 4 8 12 24 48 h
protein signal after A2
Relative protein signal NlucP / Fluc
16CAG
98CAG
NTC 0,5 1 2 3 4 8 12 24 48 h
(d)
RNA level after siHTT
Relative RNA level NlucP / Fluc
16CAG
98CAG
NTC 0,5 1 2 3 4 8 12 24 48 h
protein signal after siHTT
Relative protein signal NlucP / Fluc
16CAG
98CAG
NTC 0,5 1 2 3 4 8 12 24 48 h
(e)
Relative RNA level NlucP / GAPDH (% for each fraction)
98CAG NTC
98CAG A2
40S 60S 80S P P1 P2 P3 P4 P5 P6 P7 P8 P9
98CAG A2/NTC
difference in RNA relative level NlucP / GAPDH for A2 vs. NTC
40S 60S 80S P P1 P2 P3 P4 P5 P6 P7 P8 P9
Relative RNA level NlucP / GAPDH (% for each fraction)
16CAG NTC
16CAG A2
40S 60S 80S P P1 P2 P3 P4 P5 P6 P7 P8 P9
16CAG A2/NTC
difference in RNA relative level NlucP / GAPDH for A2 vs. NTC
40S 60S 80S P P1 P2 P3 P4 P5 P6 P7 P8 P9
(f)
98CAG 60 min.
A2
NTC
33 A
148 A
278 A
Fluorescence unit [FU]
Size [bp]
98CAG 80 min.
A2
NTC
27 A
73 A
124 A
260 A
Fluorescence unit [FU]
Size [bp]
98CAG 240 min.
A2
NTC
23 A
74 A
243 A
Fluorescence unit [FU]
Size [bp]
16CAG 60 min.
A2
NTC
150 A
284 A
Fluorescence unit [FU]
Size [bp]
16CAG 80 min.
A2
NTC
123 A
275 A
Fluorescence unit [FU]
Size [bp]
16CAG 240 min.
A2
NTC
25 A
77 A
237 A
Fluorescence unit [FU]
Size [bp]

◂**Fig. 4** Mechanistic details of A2 activity in stable cell lines with inducible expression of the *HTT* fragment. **a** Constructs used to generate Flp-In T-REx-293 cell lines with two-directional, inducible expression of the *HTT* fragment (exon 1 with 16 or 98 CAG repeats) fused with *NlucP*. *Fluc* expression was used as a reference. **b** Timeline of the experiments presented in this figure. Specific treatment and cell lysis time points are indicated. **c**, **d** Results of RT-qPCR (upper panels) and dual-luciferase assay (lower panels) to determine the *HTT-NlucP* transcript level and HTT-NlucP protein signal, respectively, after transfection of the 16CAG and 98CAG cell lines with 100 nM A2 (**c**) or siHTT (**d**) at the indicated time points. The results were normalized to the mRNA level/protein signal of Fluc in the same sample and are shown as the relative expression level/relative protein signal of HTT-NlucP in cells transfected with 100 nM control siRNA (NTC, siRLuc). $n=3$. **e** Results of RT-qPCR to assess *HTT-NlucP* expression levels in the indicated fractions containing ribosomal subunits (40S and 60S), the 80S monosome and polysomes (P-P9) after transfection of the 98CAG (upper panels) or 16 CAG (lower panels) cell lines with 100 nM A2 or control siRNA (NTC, siRLuc) at 3 h after induction. Data from each fraction were normalized to *GAPDH* expression and are presented as the % of *HTT-NlucP* expression in which 100% is the sum of the obtained values for all fractions. Graphs in the right panels show data with values calculated as the % difference in values obtained for separated fractions for control siRNA vs. A2. $n=3$. The data for **c**–**e** were analyzed using two-way ANOVA. **f** Analysis of the poly(A) tail length of the *HTT-NlucP* transcript in the 98CAG (upper panels) or 16CAG (lower panels) cell lines at the indicated time points (60, 80, 120 min) after transfection with 100 nM A2 or control siRNA (NTC, siRLuc). Estimated poly(A) tail lengths are indicated. The experiment was repeated ($n=2$), and similar results were obtained. A_0—peak obtained with reporter-specific primers to amplify a region upstream of the polyadenylation site. *An internal standard peak (1500 bp upper marker)

AGO2 (Fig. 5b). Based on previous experiments (Fig. 4c), we selected early time point of 3 h after the induction of *HTT-NlucP* expression (Fig. 5a) which we found suitable for analysis of the details of A2 activity. As in previous experiments, in this time point in the 98CAG cell line (Fig. 4c, d), both A2 and siHTT repressed *HTT-NlucP* expression up to 50% of control level (Fig. 5c). As expected for the AGO2mut and AGO2del cell lines transfected with typical siRNA, siHTT, repression of *HTT-NlucP* was completely abolished at both the transcript and protein levels and could be restored after WT AGO2 overexpression (Fig. 5c, blue bars). In contrast, after A2 transfection into AGO2mut and AGO2del cell lines, efficient lowering of both, transcript and protein levels of HTT-NlucP was achieved (Fig. 5c, orange bars). Additionally, we observed no substantial effects of WT AGO2 overexpression on A2 and siHTT activities. Together, our results show that A2-mediated downregulation of *HTT* reporter expression is mostly independent of AGO2-mediated slicer activity but is rather a consequence of transcript deadenylation and translation inhibition. Moreover, considering canonical miRNA-related mechanisms, our observations suggest that other AGO proteins (AGO1, AGO3, AGO4) and their respective miRISCs are sufficient in mutant transcript repression caused by A2, in the absence of AGO2.

Discussion

Normal and mutant alleles of polyQ diseases-related genes show varied susceptibility to regulation by art-miRNAs

Art-miRNAs were designed to target mutation site, i.e. expanded CAG repeat tract, in several transcripts implicated in polyQ diseases [25, 26]. The main rationale behind such design was to generate the universal treatment for these rare disorders. However, during research we and others (Supplementary Table 1) observed varied susceptibility of targeted transcripts to regulation by art-miRNA (Fig. 1a).

The common feature of polyQ diseases-related mRNAs is their rather low cellular level, including the brain regions mostly affected in polyQ diseases [53]. We confirmed the relatively low expression of four selected genes in patient-derived fibroblasts (*HTT*, *ATN1*, *ATXN3* and *ATXN7*), however, we also found substantial differences between the quantities of these mRNAs (Fig. 1d). In previously published studies, turnover rate and cellular abundance of transcripts were found to influence the efficiency of downregulation mediated by RNAi [54, 55]. Therefore, these factors might contribute to the observed differences in A2 activity for various transcripts, i.e., higher efficiency of *ATN1* silencing may result, at least partially, from higher expression level of this gene.

In addition to the cellular factors, we looked at polyQ disease-related transcripts and noticed that they differ in their arrangement of specific regions as well as the location of the repeat tract within the ORF (Fig. 2a). To address specific questions concerning the impact of CAG repeat tract localization on differences in allele-selective silencing by art-miRNAs, we developed cellular models with exogenous expression of the designed constructs (Fig. 2b). Silencing of exogenes expression may differ from that of endogenes, as higher expression levels of exogenes are obtained, and all additional sequences that could affect silencing efficiency are only present in endogenes. Nevertheless, we observed clear tendencies in the potency of art-miRNAs depending on the location of their targeted site and its flanking sequences (Fig. 2c). Clearly, CAG repeat tract length also determines art-miRNA activity, as large differences were observed for normal and mutant alleles silencing. Moreover, we revealed that it is crucial that the targeted sequence is present in the ORF, as the presence of the targeted tract in the 3′UTR caused efficient silencing of both normal and mutant alleles (Fig. 2c). Possibly this preference is caused by the lack of ongoing translation in 3′UTRs, in contrast to ORF regions where ribosomes interfere with the miRISC complexes [56]. Additionally, unique features of each transcript and the context of the targeted sequence can affect

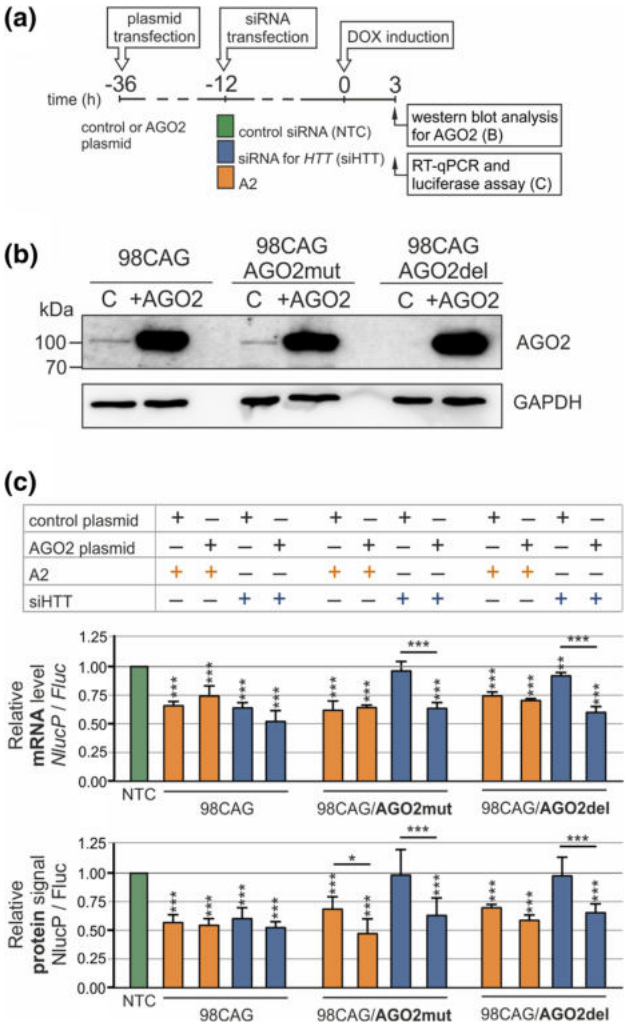


Fig. 5 Verification of the involvement of AGO2 in A2 activity. **a** Timeline for the experiments presented in this figure. Specific treatment and cell lysis time points are indicated. **b** Western blot analysis of AGO2 protein levels in 98CAG Flp-In T-REx-293 stable cell lines (*98CAG* standard cell line, *AGO2del* deletion of endogenous AGO2, *AGO2mut* abolished catalytic activity of slicer domain). *+AGO2* cell lines after transfection with AGO2 WT plasmid to rescue protein. **c** Results of RT-qPCR and luciferase assays to detect *HTT-NlucP* mRNA and protein levels, respectively, after transfection of the 98CAG cell lines (including AGO2mut and AGO2del) with 100 nM A2 or siHTT as well as the indicated plasmids. Data were normalized to *Fluc* expression levels in the same sample and *HTT-NlucP* expression levels after transfection with 100 nM control siRNA (NTC, siRLuc). The data were analyzed using two-way ANOVA (with Bonferroni multiple comparisons test among a set of samples for each of the cell lines). $n = 3$

efficiency and allele-selectivity of art-miRNAs. One such factor may be the structure formed by the CAG repeat tract, the stability of which was shown *in vitro* to be dependent on the flanking sequence (reviewed in [57]). Moreover, additional factors, like varying distance of the targeted site from STOP codon or from 3′ and 5′-ends of transcript (Fig. 2a), could contribute to diversity in efficiency of polyQ diseases-related genes silencing by A2 [58, 59].

Art-miRNAs activate events of mRNA deadenylation and translation inhibition

MiRNA-mediated regulation is known to occur in many ways depending on the activating miRNA and targeted mRNA, which can affect each other through multiple miRISC components [60, 61]. As the repertoire of miRISC proteins activate various cellular processes, the detailed analysis of a particular gene silencing mechanism is complex [3, 62]. Briefly, this silencing mechanism involves AGO-mediated recruitment of the GW182/TNRC6 protein family [51, 63], followed by subsequent binding of poly(A)-binding protein (PABPC), mRNA deadenylase complexes PAN2-PAN3 and CCR4-NOT, catalyzing deadenylation of the mRNA target, eventually leading to target decapping and transcript degradation ([64–67], reviewed in [2]). We show that art-miRNAs targeting expanded CAG repeats in ORF regions cause translation inhibition (Fig. 4e, top panel) and activate rapid mRNA deadenylation (Fig. 4f, top panel), similarly to classical miRNA pathway. Moreover, mRNA deadenylation and translational repression in typical miRISC-mediated gene silencing were also shown to be interconnected when AGO-miRNAs bound 3′UTRs (reviewed in [3, 62]). Nevertheless, the vast majority of endogenous miRNAs target 3′UTR sequences and cause mRNA decay (estimated at 66–90%) that is directly responsible for the protein downregulation [62, 68, 69]. In this study, we initially observed stronger lowering of protein level, than mRNA level (Fig. 4c), suggesting that for a pool of transcripts targeted by art-miRNAs in ORF region, translation was inhibited without activation of mRNA decay. However, when we performed detailed poly(A) tail length analysis after A2 art-miRNA treatment, deadenylation was observed already at the earliest time point analyzed after transgene induction (60 min) (Fig. 4f, top panel). On the other hand, our data obtained from polysome profiling suggested that A2-activated translational repression occurs also very rapidly, 3 h after transgene induction (Fig. 4e, top panel). Therefore, we cannot exclude that deadenylation preceded or was concurrent with translation inhibition. Indeed, in some cases, translational inhibition was shown to precede poly(A) tail shortening and mRNA decay. Such cases were: a study performed in HeLa cells [8] and a study using a Drosophila S2 cell-based controllable expression system [70] where miRNA-targeted sequences were localized in the 3′UTR. Here, to understand the investigated mechanism, it was also crucial to determine if ribosome complexes are formed on targeted transcripts as the result of art-miRNA activity. Our data obtained from polysome profiling suggest that A2-activated translational repression can occur at the

elongation step and/or at the initiation step (Fig. 4e). This conclusion is supported by the comparison of the shift of *HTT* transcript, observed after A2 treatment, with the profiles of mRNA distributions in monosome and polysome fractions characteristic for global inhibition of translation at the initiation or elongation step [71, 72].

Transcript-dependent factors, affecting the activation of deadenylation and translation inhibition processes, may also contribute to differences in the effectiveness of art-miRNA in different models of polyQ diseases. According to translation-dependent closed-loop model [73, 74], it can be assumed that, although CAG repeats are in a very large distance from the poly(A) tail (especially for *HTT* transcript ~ 13 kb), art-miRNA-bound miRISCs are in close proximity to the poly(A) tails of polyQ disease-related mRNAs in cells (Fig. 6). In this case, miRISC-mediated translational repression can also occur through recruitment of the RNA helicase DDX6, which acts as both a translational inhibitor and decapping activator [75–77]. Moreover, it was also shown that DDX6 can act by displacing the eukaryotic translation initiation factors eIF4A-I and eIF4A-II from the targeted transcripts, thereby preventing translation initiation [10, 78, 79]. These processes are dependent on features of 3′UTR region of targeted transcript, as various cis- and trans-acting elements in specific 3′UTRs were found to influence miRNA-mediated gene expression regulation [80].

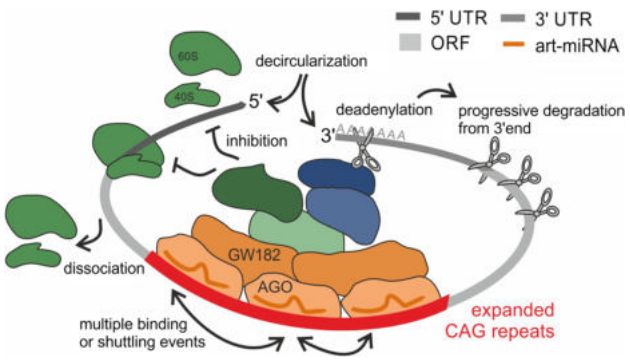


Fig. 6 Model of art-miRNA activity targeting transcripts containing expanded CAG repeat tracts within the ORF region. Art-miRNA loaded into AGO binds to mutant CAG repeat tract with a mismatch formed in the central region of this interaction. Multiple binding or shuttling of the AGO protein results in the formation of the miRISC, which affects translation by the inhibition of its initiation or early elongation. Shortening of the poly(A) tail is activated for a pool of targeted transcripts that leads to the subsequent degradation of mRNA

The art-miRNA mechanism as a model for the AGO-dependent cooperative activities of miRNAs within ORF regions

It is known that miRNAs regulate genes expression mostly by recognition of sites within 3′UTR, however numerous miRNA-binding sites were revealed by global approaches also in the ORFs of human mRNAs [81, 82]. The functionalities of approximately twenty sites of this type in specific transcripts have been experimentally confirmed so far (reviewed in [83]), but precise mechanisms have not been extensively investigated. Recently, a specific type of miRNA recognition elements exclusive to ORF regions was described, and a mechanism of gene expression regulation by temporary ribosome stalling was proposed for DAPK3 kinase [84]. Based on the results of our study, we can extrapolate mechanistic details of ORF regions-targeting by miRNAs, especially when multiple binding sites are present. Indeed, for miRNA-based regulation within ORFs, multiple binding sites have been found frequently [83]. One such example is the regulation of the expression of a family of genes containing C_2H_2 zinc-finger domains by a group of miRNAs [85, 86]. Additionally, a general role of repeat tracts localized in ORFs in post-transcriptional regulation was suggested based on predicted interactions [86]. Overall, these data and results regarding art-miRNA activity suggest that for the efficient regulation of a gene's expression by targeting its ORF, multiple binding sites are required, resulting in cooperative action. Some mechanistic details of the cooperative activity of miRNAs were revealed, e.g., a FRET-based method was used to show that AGO2 dissociation is in kinetic competition with lateral diffusion, resulting in shuttling between adjacent target sites [87].

In general, human cells express four AGO paralogs (AGO1-4) that act to regulate miRNA-based gene expression [88]. AGO2 is the most abundant, as it accounts for ~ 70% of the total AGO pool in HEK 293T and fibroblast cells [89, 90], in which we performed most of mechanistic experiments. Previously, discrimination between the normal and mutant alleles of *HTT* mRNA by art-miRNAs was reported to be highly sensitive to the cellular pool of AGO2 and GW182 family proteins [30]. Our analysis in total AGO2 knockout and AGO2 D597A endonuclease-deficient cell lines, showed that the absence of AGO2 does not affect the observed silencing activity of exemplary A2 art-miRNA (Fig. 5c). These results suggest that other slicer-deficient AGOs (AGO1, 3 and/or 4) can act in the cooperative repression of the mutant allele. Nevertheless, we do not rule out that AGO2 is a key, most abundant miRISC core protein, but we show it may be replaced by other AGOs. These results are consistent with previous observations that the majority of human miRNAs associate with all four AGOs and do not have a preference for a particular AGO paralog [20, 91, 92]. We assume, that after art-miRNA-AGO complex binding to mRNA, the subsequent co-recruitment of the GW182/TNRC6 with other effector miRISC proteins results in allele-selective inhibition of the mutant allele. Our data suggest that art-miRNAs can be additionally recruited by slicer-deficient AGO proteins (AGO1, 3 and/or 4) to the expanded

CAG repeats (Fig. 6) what might turn out to be advantageous for the experimental therapy based on art-miRNA reagents. This is due to the previous RNA-seq and mass spectrometry analysis that clearly indicate that in brain tissue the relative abundance of AGO1, AGO3 and/or AGO4 (when related to the total AGO pool) are higher than in many other cells and tissues analyzed [89, 93]. This particularly applies to AGO1 protein as quantitative proteomic approach revealed in HEK 293T cells the following proportions of AGO proteins: ~17% AGO1, ~75% AGO2, ~6% AGO3 and ~2% AGO4, whereas similar mass spectrometry analysis conducted on mouse brain lysates showed proportions: ~35% AGO1, ~55% AGO2, ~9% AGO3 and < 1% AGO4 [89].

Art-miRNA activity in the context of other therapeutic strategies for polyQ diseases

We are now witnessing large advances in antisense oligonucleotide (ASO)- and RNAi-based strategies in clinical trials [94, 95]. The most advanced clinical trial of a causative therapy for HD involves the intrathecal delivery of RNase-H-activating ASOs targeting *HTT* (ClinicalTrials.gov Identifier: NCT03842969). The results obtained thus far are very promising, as a decrease in huntingtin in the spinocerebellar fluid was reported [96]. Nevertheless, many challenges remain to be faced in the developed therapy, among which the allele-selectivity of silencing is one of the major points [97]. Preservation of the level of the normal allele might be required, as its long-term downregulation, which would be the result of long-term treatment, could have many adverse effects [98, 99]. Additionally, in the case of HD, some reports suggest that targeting exon 1 of *HTT*, which contains the CAG tract, may be crucial to eliminate key toxic entities causing HD pathogenesis [100, 101]. For these reasons, we find an approach using art-miRNAs to be desirable. This CAG repeat-targeting strategy offers an option for the preferential silencing of several mutant alleles responsible for polyQ diseases and would be applicable to a larger group of patients than an allele-selective SNP-targeting-based approach. Moreover, due to somatic instability, mosaicism of highly expanded CAG repeats in the brain is likely a common effect largely responsible for brain-specific pathology, as shown in HD [102–104]. In this case, transcripts containing increasingly expanded CAG repeat tracts are expected to be more efficiently targeted by art-miRNAs. This assumption remains to be proven experimentally but would clearly allow preferential targeting of the RNAs which translation leads to pathogenesis.

The potential universality of targeting the mutation site in RNA was also shown for the activity of ASOs acting as translation blockers or splicing modulators that were tested for several polyQ diseases [105–107]. According to recent findings, an art-miRNA-based therapeutic strategy might not be applicable for SCA1 [108], and the development of a universal molecule for several polyQ diseases might require further research. Although more demanding than initially assumed, the applicability of this strategy in at least a few disorders remains feasible. Art-miRNAs possess additional advantages as they can be chemically modified [31, 38] or expressed from vectors [33, 109], and can include application of novel approaches for delivery to the brain [110, 111]. Interestingly, CRISPR-Cas9- and ZFP-based CAG repeat-targeting strategies were recently successfully tested in HD models [112–116]. These approaches offer an alternative solution for mutant *HTT* inhibition after the binding of specifically designed molecules to mutant DNA, but some challenges remain before their clinical testing.

In summary, our model of art-miRNAs activity show potential versatility in the miRNA-based regulation of gene expression. Although this model (Fig. 6) is based on results obtained for artificial miRNA, it contributes to a better understanding of the mechanisms of action of natural miRNAs which interact with sequences located in ORFs with adjacent multiple binding sites. These mechanisms have been harnessed to activate the therapeutically beneficial silencing of mutant genes with CAG repeat expansions, showing the great flexibility of RNAi-based mechanisms in cells.

Acknowledgements This work was supported by Grants from National Science Centre [2014/15/B/NZ1/01880 to WJK/AF, 2015/17/D/NZ5/03443 to AF, 2015/19/B/NZ2/02453 to WJK/ACie, 2015/17/N/NZ2/01916 to EK]; and Polish Ministry of Science and Higher Education [DI 2011 0278 41 to ASC and 01/KNOW2/2014, the KNOW program]. Microscopic images were obtained in the Laboratory of Subcellular Structures Analysis, IBCH PAS. The authors would like to thank former and current members of the Department for discussions and contributions to the preliminary analyses of this study.

Author contributions Conceptualization: WJK, AF, ACie and ASC. Transfections and western blot on fibroblasts: AF, ASC and DZ. DdPCR: AF and MWW. Design of luciferase-based plasmids, relevant experiments and bioinformatics analyses of transcripts features: ASC. Design of luciferase-based plasmids used for Flp-In T-REx cell line generation, luciferase assay experiments: ACie. Flp-In cell lines, RT-qPCR and poly(A) tail length assay experiments: ACie, PJ. Optimization of polysome profiling method: ACio, KDR and DZ. Polysome profiling experiments and relevant RT-qPCR: ACio. Generation of NPs and relevant smFISH: EK. SmFISH on fibroblasts: MM. AGO2 mutant and knockout cell lines: MD, MO and ACie. Manuscript writing: AF, ASC and ACie, with the input and revision from all authors.

Compliance with ethical standards

Conflict of interest 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.

References

1. Sonenberg N, Fabian MR (2012) The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. Nat Struct Mol Biol 19:586–593. https://doi.org/10.1038/nsmb.2296
2. Stroynowska-Czerwinska A, Fiszer A, Krzyzosiak WJ (2014) The panorama of miRNA-mediated mechanisms in mammalian cells. Cell Mol Life Sci 71:2253–2270. https://doi.org/10.1007/s00018-013-1551-6
3. Jonas S, Izaurralde E (2015) Towards a molecular understanding of microRNA-mediated gene silencing. Nat Rev Genet 16:421–433. https://doi.org/10.1038/nrg3965
4. Bartel DP (2018) Metazoan MicroRNAs. Cell 173:20–51. https://doi.org/10.1016/J.CELL.2018.03.006
5. Elbashir SM, Harborth J, Lendeckel W et al (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411:494–498
6. Pillai RS, Bhattacharyya SN, Artus CG et al (2005) Inhibition of translational initiation by Let-7 MicroRNA in human cells. Science 309:1573–1576. https://doi.org/10.1126/science.1115079
7. Eulalio A, Huntzinger E, Nishihara T et al (2009) Deadenylation is a widespread effect of miRNA regulation. RNA 15:21–32. https://doi.org/10.1261/rna.1399509
8. Béthune J, Artus-Revel CG, Filipowicz W (2012) Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. EMBO Rep 13:716–723. https://doi.org/10.1038/embor.2012.82
9. Huntzinger E, Kuzuoglu-Öztürk D, Braun JE et al (2013) The interactions of GW182 proteins with PABP and deadenylases are required for both translational repression and degradation of miRNA targets. Nucleic Acids Res 41:978–994. https://doi.org/10.1093/nar/gks1078
10. Meijer HA, Kong YW, Lu WT et al (2013) Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation. Science 340:82–85. https://doi.org/10.1126/science.1231197
11. Duursma AM, Kedde M, Schrier M et al (2008) miR-148 targets human DNMT3b protein coding region. RNA 14:872–877. https://doi.org/10.1261/rna.972008
12. Forman JJ, Legesse-Miller A, Coller HA (2008) A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. Proc Natl Acad Sci USA 105:14879–14884
13. Tay Y, Zhang J, Thomson AM et al (2008) MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. Nature 455:1124–1128
14. Elcheva I, Goswami S, Noubissi FK, Spiegelman VS (2009) CRD-BP protects the coding region of betaTrCP1 mRNA from miR-183-mediated degradation. Mol Cell 35:240–246. https://doi.org/10.1016/j.molcel.2009.06.007
15. Forman JJ, Coller HA (2010) The code within the code: MicroRNAs target coding regions. Cell Cycle 9:1533–1541. https://doi.org/10.4161/cc.9.8.11202
16. Jopling CL, Yi M, Lancaster AM et al (2005) Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. Science 309:1577–1581. https://doi.org/10.1126/science.1113329
17. Ørom UA, Nielsen FC, Lund AH (2008) MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. Mol Cell 30:460–471. https://doi.org/10.1016/j.molcel.2008.05.001
18. Tsai N-P, Lin Y-L, Wei L-N (2009) MicroRNA mir-346 targets the 5'-untranslated region of receptor-interacting protein 140 (RIP140) mRNA and up-regulates its protein expression. Biochem J 424:411–418. https://doi.org/10.1042/BJ20090915
19. Gu S, Jin L, Zhang F et al (2009) The biological basis for microRNA target restriction to the 3′ untranslated region in mammalian mRNAs. Nat Struct Mol Biol 16:144–150. https://doi.org/10.1038/nsmb.1552.The
20. Broderick JA, Salomon WE, Ryder SP et al (2011) Argonaute protein identity and pairing geometry determine cooperativity in mammalian RNA silencing. RNA 17:1858–1869. https://doi.org/10.1261/rna.2778911.promotes
21. Grimson A, Farh KK-H, Johnston WK et al (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell 27:91–105
22. Saetrom P, Heale BSE, Snøve O et al (2007) Distance constraints between microRNA target sites dictate efficacy and cooperativity. Nucleic Acids Res 35:2333–2342
23. Orr HT (2012) Polyglutamine neurodegeneration: expanded glutamines enhance native functions. Curr Opin Genet Dev 22:251–255
24. Stoyas CA, La Spada AR (2018) The CAG–polyglutamine repeat diseases: a clinical, molecular, genetic, and pathophysiologic nosology. In: Handbook of clinical neurology. pp 143–170. https://doi.org/10.1016/B978-0-444-63233-3.00011-7
25. Matsui M, Corey DR (2012) Allele-selective inhibition of trinucleotide repeat genes. Drug Discov Today 17:443–450
26. Fiszer A, Krzyzosiak WJ (2014) Oligonucleotide-based strategies to combat polyglutamine diseases. Nucleic Acids Res 42:6787–6810. https://doi.org/10.1093/nar/gku385
27. Hu J, Liu J, Corey DR (2010) Allele-selective inhibition of huntingtin expression by switching to an miRNA-like RNAi mechanism. Chem Biol 17:1183–1188
28. Fiszer A, Mykowska A, Krzyzosiak WJ (2011) Inhibition of mutant huntingtin expression by RNA duplex targeting expanded CAG repeats. Nucleic Acids Res 39:5578–5585
29. Hu J, Gagnon KT, Liu J et al (2011) Allele-selective inhibition of ataxin-3 (ATX3) expression by antisense oligomers and duplex RNAs. Biol Chem 392:315–325
30. Hu J, Liu J, Yu D et al (2012) Mechanism of allele-selective inhibition of huntingtin expression by duplex RNAs that target CAG repeats: function through the RNAi pathway. Nucleic Acids Res 40:11270–11280. https://doi.org/10.1093/nar/gks907
31. Yu D, Pendergraff H, Liu J et al (2012) Single-stranded RNAs use RNAi to potently and allele-selectively inhibit mutant huntingtin expression. Cell 150:895–908. https://doi.org/10.1016/j.cell.2012.08.002
32. Aiba Y, Hu J, Liu J et al (2013) Allele-selective inhibition of huntingtin and ataxin-3 expression by RNA duplexes containing unlocked nucleic acid (UNA) substitutions. Biochemistry 52:9329–9338. https://doi.org/10.1021/bi4014209
33. Fiszer A, Olejniczak M, Galka-Marciniak P et al (2013) Self-duplexing CUG repeats selectively inhibit mutant huntingtin expression. Nucleic Acids Res 41:10426–10437. https://doi.org/10.1093/nar/gkt825

34. Liu J, Pendergraff H, Narayanannair KJ et al (2013) RNA duplexes with abasic substitutions are potent and allele-selective inhibitors of huntingtin and ataxin-3 expression. Nucleic Acids Res 41:8788–8801. https://doi.org/10.1093/nar/gkt594
35. Liu J, Yu D, Aiba Y et al (2013) ss-siRNAs allele selectively inhibit ataxin-3 expression: multiple mechanisms for an alternative gene silencing strategy. Nucleic Acids Res 41:9570–9583. https://doi.org/10.1093/nar/gkt693
36. Hu J, Liu J, Yu D et al (2014) Exploring the effect of sequence length and composition on allele-selective inhibition of human huntingtin expression by single-stranded silencing RNAs. Nucleic Acid Ther. https://doi.org/10.1089/nat.2013.0476
37. Hu J, Liu J, Narayanannair KJ et al (2014) Allele-selective inhibition of mutant Atrophin-1 expression by duplex and single-stranded RNAs. Biochemistry 53:4510–4518. https://doi.org/10.1021/bi500610r
38. Fiszer A, Ellison-Klimontowicz ME, Krzyzosiak WJ (2016) Silencing of genes responsible for polyQ diseases using chemically modified single-stranded siRNAs. Acta Biochim Pol. https://doi.org/10.18388/abp.2016_1336
39. Fiszer A, Wroblewska J, Nowak B, Krzyzosiak W (2016) Mutant CAG repeats effectively targeted by RNA interference in SCA7 cells. Genes (Basel) 7:132. https://doi.org/10.3390/genes7120132
40. Urbanek MO, Fiszer A, Krzyzosiak WJ (2017) Reduction of Huntington's disease RNA foci by CAG repeat-targeting reagents. Front Cell Neurosci 11:82. https://doi.org/10.3389/fncel.2017.00082
41. Takahashi N, Sasagawa N, Suzuki K, Ishiura S (1999) Synthesis of long trinucleotide repeats in vitro. Neurosci Lett 262:45–48. https://doi.org/10.1016/s0304-3940(99)00031-2
42. Figura G, Koscianska E, Krzyzosiak W (2015) In vitro expansion of CAG, CAA, and mixed CAG/CAA repeats. Int J Mol Sci 16:18741–18751. https://doi.org/10.3390/ijms160818741
43. Sammarco MC, Grabczyk E (2005) A series of bidirectional tetracycline-inducible promoters provides coordinated protein expression. Anal Biochem 346:210–216. https://doi.org/10.1016/j.ab.2005.08.033
44. Suzuki K, Bose P, Leong-Quong RY et al (2010) REAP: a two minute cell fractionation method. BMC Res Notes 3:294. https://doi.org/10.1186/1756-0500-3-294
45. Faye MD, Graber TE, Holcik M (2014) Assessment of selective mRNA translation in mammalian cells by polysome profiling. J Vis Exp. https://doi.org/10.3791/52295
46. Blair JD, Hockemeyer D, Doudna JA et al (2017) Widespread translational remodeling during human neuronal differentiation. Cell Rep 21:2005–2016. https://doi.org/10.1016/j.celrep.2017.10.095
47. Kusov YY, Shatirishvili G, Dzagurov G, Gauss-Müller V (2001) A new G-tailing method for the determination of the poly(A) tail length applied to hepatitis A virus RNA. Nucleic Acids Res 29:E57–E67. https://doi.org/10.1093/nar/29.12.e57
48. Meister G, Landthaler M, Patkaniowska A et al (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell 15:185–197. https://doi.org/10.1016/j.molcel.2004.07.007
49. Gu S, Jin L, Zhang F et al (2009) Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. Nat Struct Mol Biol 16:144–150. https://doi.org/10.1038/nsmb.1552
50. Didiot M-C, Ferguson CM, Ly S et al (2018) Nuclear localization of huntingtin mRNA is specific to cells of neuronal origin. Cell Rep 24:2553–2560.e5. https://doi.org/10.1016/j.celrep.2018.07.106
51. Liu J, Carmell MA, Rivas FV et al (2004) Argonaute2 is the catalytic engine of mammalian RNAi. Science (80-) 305:1437–1441
52. Rivas FV, Tolia NH, Song JJ et al (2005) Purified Argonaute2 and an siRNA form recombinant human RISC. Nat Struct Mol Biol 12:340–349
53. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL et al (2012) An anatomically comprehensive atlas of the adult human brain transcriptome. Nature 489:391–399. https://doi.org/10.1038/nature11405
54. Larsson E, Sander C, Marks D (2010) mRNA turnover rate limits siRNA and microRNA efficacy. Mol Syst Biol 6:433. https://doi.org/10.1038/msb.2010.89
55. Hong SW, Jiang Y, Kim S et al (2014) Target gene abundance contributes to the efficiency of siRNA-mediated gene silencing. Nucleic Acid Ther 24:192. https://doi.org/10.1089/NAT.2013.0466
56. Gu W, Xu Y, Xie X et al (2014) The role of RNA structure at 5′ untranslated region in microRNA-mediated gene regulation. RNA 20:1369–1375. https://doi.org/10.1261/rna.044792.114
57. Ciesiolka A, Jazurek M, Drazkowska K, Krzyzosiak WJ (2017) Structural characteristics of simple RNA repeats associated with disease and their deleterious protein interactions. Front Cell Neurosci 11:97. https://doi.org/10.3389/fncel.2017.00097
58. Fang Z, Rajewsky N (2011) The impact of miRNA target sites in coding sequences and in 3′UTRs. PLoS ONE 6:e18067. https://doi.org/10.1371/journal.pone.0018067
59. Duchaine TF, Fabian MR (2019) Mechanistic insights into MicroRNA-mediated gene silencing. Cold Spring Harb Perspect Biol 11:a032771. https://doi.org/10.1101/cshperspect.a032771
60. Kalantari R, Hicks JA, Li L et al (2016) Stable association of RNAi machinery is conserved between the cytoplasm and nucleus of human cells. RNA. https://doi.org/10.1261/rna.056499.116
61. Trabucchi M (2019) Subcellular Heterogeneity of the microRNA Machinery. Trends Genet 35:15–28. https://doi.org/10.1016/J.TIG.2018.10.006
62. Gebert LFR, MacRae IJ (2019) Regulation of microRNA function in animals. Nat Rev Mol Cell Biol 20:21–37. https://doi.org/10.1038/s41580-018-0045-7
63. Rehwinkel J, Behm-Ansmant I, Gatfield D, Izaurralde E (2005) A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. RNA 11:1640–1647. https://doi.org/10.1261/rna.2191905
64. Behm-Ansmant I, Rehwinkel J, Doerks T et al (2006) mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. Genes Dev 20:1885–1898. https://doi.org/10.1101/gad.1424106
65. Braun JE, Huntzinger E, Fauser M, Izaurralde E (2011) GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. Mol Cell 44:120–133. https://doi.org/10.1016/j.molcel.2011.09.007
66. Chekulaeva M, Mathys H, Zipprich JT et al (2011) miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. Nat Struct Mol Biol 18:1218–1226. https://doi.org/10.1038/nsmb.2166
67. Fabian MR, Cieplak MK, Frank F et al (2011) miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. Nat Struct Mol Biol 18:1211–1217. https://doi.org/10.1038/nsmb.2149
68. Guo H, Ingolia NT, Weissman JS, Bartel DP (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 466:835–840
69. Eichhorn SWW, Guo H, McGeary SEE et al (2014) mRNA destabilization is the dominant effect of mammalian MicroRNAs by

the time substantial repression ensues. Mol Cell 56:104–115. https://doi.org/10.1016/j.molcel.2014.08.028
70. Djuranovic S, Nahvi A, Green R (2012) miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. Science (80-) 336:237–240. https://doi.org/10.1126/science.1215691
71. Li CH, Ohn T, Ivanov P et al (2010) eIF5A promotes translation elongation, polysome disassembly and stress granule assembly. PLoS ONE 5:e9942. https://doi.org/10.1371/journal.pone.0009942
72. Chassé H, Boulben S, Costache V et al (2017) Analysis of translation using polysome profiling. Nucleic Acids Res 45:gkw07. https://doi.org/10.1093/nar/gkw907
73. Vicens Q, Kieft JS, Rissland OS (2018) Revisiting the closed-loop model and the nature of mRNA 5′–3′ communication. Mol Cell 72:805–812
74. Fakim H, Fabian MR (2019) Communication is key: 5′–3′ interactions that regulate mRNA translation and turnover. Adv Exp Med Biol 1203:149–164. https://doi.org/10.1007/978-3-030-31434-7_6
75. Mathys H, Basquin J, Ozgur S et al (2014) Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in MicroRNA repression. Mol Cell 54:751–765. https://doi.org/10.1016/j.molcel.2014.03.036
76. Rouya C, Siddiqui N, Morita M et al (2014) Human DDX6 effects miRNA-mediated gene silencing via direct binding to CNOT1. RNA 20:1398–1409. https://doi.org/10.1261/rna.045302.114
77. Kamenska A, Simpson C, Vindry C et al (2016) The DDX6-4E-T interaction mediates translational repression and P-body assembly. Nucleic Acids Res 44:6318–6334. https://doi.org/10.1093/nar/gkw565
78. Fukao A, Mishima Y, Takizawa N et al (2014) MicroRNAs trigger dissociation of eIF4AI and eIF4AII from target mRNAs in humans. Mol Cell 56:79–89. https://doi.org/10.1016/j.molcel.2014.09.005
79. Fukaya T, Iwakawa H-O, Tomari Y (2014) MicroRNAs block assembly of eIF4F translation initiation complex in Drosophila. Mol Cell 56:67–78. https://doi.org/10.1016/j.molcel.2014.09.004
80. Matoulkova E, Michalova E, Vojtesek B, Hrstka R (2012) The role of the 3′ untranslated region in post-transcriptional regulation of protein expression in mammalian cells. RNA Biol 9:563–576. https://doi.org/10.4161/rna.20231
81. Hafner M, Landthaler M, Burger L et al (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141:129–141. https://doi.org/10.1016/j.cell.2010.03.009
82. Helwak A, Kudla G, Dudnakova T, Tollervey D (2013) Mapping the human miRNA interactome by CLASH reveals frequent non-canonical binding. Cell 153:654–665. https://doi.org/10.1016/j.cell.2013.03.043
83. Brümmer A, Hausser J (2014) MicroRNA binding sites in the coding region of mRNAs: extending the repertoire of post-transcriptional gene regulation. BioEssays 36:617–626. https://doi.org/10.1002/bies.201300104
84. Zhang K, Zhang X, Cai Z et al (2018) A novel class of microRNA-recognition elements that function only within open reading frames. Nat Struct Mol Biol 25:1019–1027. https://doi.org/10.1038/s41594-018-0136-3
85. Huang S, Wu S, Ding J et al (2010) MicroRNA-181a modulates gene expression of zinc finger family members by directly targeting their coding regions. Nucleic Acids Res 38:7211–7218. https://doi.org/10.1093/nar/gkq564
86. Schnall-Levin M, Rissland OS, Johnston WK et al (2011) Unusually effective microRNA targeting within repeat-rich coding regions of mammalian mRNAs. Genome Res 21:1395–1403. https://doi.org/10.1101/gr.121210.111
87. Chandradoss SD, Schirle NT, Szczepaniak M et al (2015) A dynamic search process underlies MicroRNA targeting. Cell 162:96–107. https://doi.org/10.1016/j.cell.2015.06.032
88. Dueck A, Meister G (2014) Assembly and function of small RNA—Argonaute protein complexes. Biol Chem. https://doi.org/10.1515/hsz-2014-0116
89. Hauptmann J, Schraivogel D, Bruckmann A et al (2015) Biochemical isolation of Argonaute protein complexes by Ago-APP. Proc Natl Acad Sci USA 112:11841–11845. https://doi.org/10.1073/pnas.1506116112
90. Völler D, Linck L, Bruckmann A et al (2016) Argonaute family protein expression in normal tissue and cancer entities. PLoS ONE 11:e0161165. https://doi.org/10.1371/journal.pone.0161165
91. Dueck A, Ziegler C, Eichner A et al (2012) microRNAs associated with the different human Argonaute proteins. Nucleic Acids Res 40:9850–9862. https://doi.org/10.1093/nar/gks705
92. Wang D, Zhang Z, O'Loughlin E et al (2012) Quantitative functions of Argonaute proteins in mammalian development. Genes Dev 26:693–704. https://doi.org/10.1101/gad.182758.111
93. Uhlen M, Fagerberg L, Hallstrom BM et al (2015) Tissue-based map of the human proteome. Science (80-) 347:1260419–1260419. https://doi.org/10.1126/science.1260419
94. Shen X, Corey DR (2018) Chemistry, mechanism and clinical status of antisense oligonucleotides and duplex RNAs. Nucleic Acids Res 46:1584–1600. https://doi.org/10.1093/nar/gkx1239
95. Setten RL, Rossi JJ, Han S (2019) The current state and future directions of RNAi-based therapeutics. Nat Rev Drug Discov 18:421–446. https://doi.org/10.1038/s41573-019-0017-4
96. Tabrizi SJ, Leavitt BR, Landwehrmeyer GB et al (2019) Targeting huntingtin expression in patients with Huntington's disease. N Engl J Med. https://doi.org/10.1056/NEJMoa1900907
97. 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
98. Saudou F, Humbert S (2016) The biology of huntingtin. Neuron 89:910–926. https://doi.org/10.1016/j.neuron.2016.02.003
99. Liu J-P, Zeitlin SO (2017) Is huntingtin dispensable in the adult brain? J Huntingtons Dis 6:1–17. https://doi.org/10.3233/JHD-170235
100. Neueder A, Landles C, Ghosh R et al (2017) The pathogenic exon 1 HTT protein is produced by incomplete splicing in Huntington's disease patients. Sci Rep 7:1307. https://doi.org/10.1038/s41598-017-01510-z
101. Franich NR, Hickey MA, Zhu C et al (2019) Phenotype onset in Huntington's disease knock-in mice is correlated with the incomplete splicing of the mutant huntingtin gene. J Neurosci Res 97:jnr24493. https://doi.org/10.1002/jnr.24493
102. Telenius H, Kremer B, Goldberg YP et al (1994) Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. Nat Genet 6:409–414. https://doi.org/10.1038/ng0494-409
103. Kennedy L, Evans E, Chen C-M et al (2003) Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. Hum Mol Genet 12:3359–3367. https://doi.org/10.1093/hmg/ddg352
104. Wright GEB, Collins JA, Kay C et al (2019) Length of uninterrupted CAG, independent of polyglutamine size, results in increased somatic instability, hastening onset of huntington disease. Am J Hum Genet 104:1116–1126. https://doi.org/10.1016/j.ajhg.2019.04.007
105. Hu J, Matsui M, Gagnon KT 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. https://doi.org/10.1038/nbt.1539
106. Datson NA, González-Barriga A, Kourkouta E et al (2017) The expanded CAG repeat in the huntingtin gene as target for therapeutic RNA modulation throughout the HD mouse brain. PLoS ONE 12:e0171127. https://doi.org/10.1371/journal.pone.0171127
107. Kourkouta E, Weij R, González-Barriga A et al (2019) Suppression of mutant protein expression in SCA3 and SCA1 mice using a CAG repeat-targeting antisense oligonucleotide. Mol Ther Nucleic Acids 17:601–614. https://doi.org/10.1016/j.omtn.2019.07.004
108. Hu J, Corey DR (2019) Limits of using oligonucleotides for allele-selective inhibition at trinucleotide repeat sequences–targeting the CAG repeat within ataxin-1. Nucleosides Nucleotides Nucleic Acids. https://doi.org/10.1080/15257770.2019.1671592
109. Kotowska-Zimmer A, Ostrovska Y, Olejniczak M (2020) Universal RNAi triggers for the specific inhibition of mutant huntingtin, atrophin-1, ataxin-3, and ataxin-7 expression. Mol Ther Nucleic Acids 19:562–571. https://doi.org/10.1016/j.omtn.2019.12.012
110. Chan KY, Jang MJ, Yoo BB et al (2017) Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. Nat Neurosci 20:1172–1179. https://doi.org/10.1038/nn.4593
111. Alterman JF, Godinho BMDC, Hassler MR et al (2019) A divalent siRNA chemical scaffold for potent and sustained modulation of gene expression throughout the central nervous system. Nat Biotechnol 37:884–894. https://doi.org/10.1038/s41587-019-0205-0
112. 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
113. 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
114. Zeitler B, Froelich S, Marlen K et al (2019) Allele-selective transcriptional repression of mutant HTT for the treatment of Huntington's disease. Nat Med. https://doi.org/10.1038/s41591-019-0478-3
115. 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
116. Dabrowska M, Ciolak A, Kozlowska E et al (2020) Generation of new isogenic models of Huntington's disease using CRISPR-Cas9 technology. Int J Mol Sci 21:1854. https://doi.org/10.3390/ijms21051854

SUPPLEMENTARY DATA

SUPPLEMENTARY TEXT

Features of *HTT* and *ATXN3* transcripts which may affect allele-selective regulation by art-miRNAs

The results of luciferase assay (Fig. 2C) suggest that within sequence of *HTT* transcript there are "positive factors" for allele-selective silencing (decreasing silencing of normal allele and/or increasing silencing of mutated allele), whereas within *ATXN3* mRNA some "negative factors" can be present. We analyzed *HTT* and *ATXN3* transcripts with the use of bioinformatic approach, i.e. RegRNA 2.0 web server for identifying functional RNA motifs and binding sites [1]. RegRNA for *ATXN3* transcript sequence around CAG repeat tract did not show any potential binding sites for molecules. For *HTT* mRNA we identified miR-3960 binding site with near fully complementary binding, including lack of mismatches in seed region and single mismatch in central region. Nevertheless, experimental validation showed that miRNA-3960 did not improve significantly silencing efficiency of art-miRNAs (data not shown).

Moreover, we checked the frequency of codons upstream of CAG repeat tract within tested transcripts (*HTT*, *ATXN3* and *Rluc*) (Fig. S3). The analysis is based on observation that translation inhibition efficiency by miRNAs targeting ORF regions may dependent on the presence of rare codons from the translation start to miRNA binding site [2]. Slowing down the ribosomes on the transcript, caused by the translation of rare codons, was shown to increase downregulation efficiency by miRNAs targeting ORF regions [2]. Three analyzed transcripts contained codons with average codon usage values similar to optimal codons used by Gu et al. However, for comparison of *HTT* and *ATXN3* transcripts we did not observed significant differences in their average codon usage values for regions upstream to CAG repeats. Therefore these analyses do not justify observed differences in allele-selectivity in art-miRNAs activity in HD and SCA3 models.

SUPPLEMENTARY MATERIALS AND METHODS

Immunocytochemistry (ICC)

NP cells were fixed in 4% PFA, permeabilized with 0.5% Tween and blocked in 1% bovine serum albumin (all in 1xPBS), followed by incubation with primary antibodies and fluorescent-dye conjugated secondary antibodies (after wash in PBS), listed in Supplementary Table 3. DAPI was used for nuclei staining. Images were captured with Leica DMI6000 microscope as described in Methods section.

Details on generation of pcDNA5/FRT/TO-based constructs for inducible and stable expression of CAG repeat target sequence

Each intermediate stage of cloning and the final modified pcDNA5/FRT/TO-based plasmids were confirmed by Sanger DNA sequencing. All PCR fragments used for cloning were obtained with Q5 High-Fidelity DNA Polymerase (New England BioLabs) according to manufacturer's instructions. Sequences of DNA oligonucleotides used for cloning are given in Supplementary Table 7. The following steps were performed:

1) *Fluc* sequence was PCR amplified from pmirGLO plasmid (Promega) using primers (1/2). Next, the *Fluc* fragment was digested with KpnI/BamHI restriction enzymes, and cloned between KpnI/BamHI sites in pcDNA3.1(+) plasmid (Invitrogen) to obtain Fluc_pcDNA3.1(+) plasmid.

2) SV40pA was PCR amplified from pNL1.2 plasmid (Promega) using primers (3/4). Next, the SV40pA fragment was digested with BglII/BamHI restriction enzymes, and cloned, in a correct orientation, between BamHI site in Fluc_pcDNA3.1(+) plasmid to obtain Fluc_SV40pA pcDNA3.1(+) plasmid.

3) Fluc_SV40pA_pcDNA3.1(+) plasmid was digested with KpnI/BamHI restriction enzymes to obtain Fluc_SV40pA fragment. Next, the Fluc_SV40pA insert was cloned, in reverse orientation, between BglII/KpnI sites in pcDNA5/FRT/TO plasmid (Invitrogen) to obtain SV40pA_Fluc pcDNA5/FRT/TO plasmid.

4) The generation of bidirectional tetracycline-inducible BI-16 promoter, a derivate of the original CMV promoter duplicated back-to-back was done as in [3] with minor modifications. Asymmetric fragments 1 and 2 of PCMV(2x-TetO2) promoter from pcDNA5/FRT/TO were amplified by PCR and then ligated together. Fragment 1 was amplified using primers (5/6), while fragment 2 was amplified using primers (7/8). PCR products were cut with XbaI, mixed and ligated using T4 DNA ligase to make an asymmetric inverted repeat. The intermediated-length products of these ligations, representing the heterodimers were purified from agarose gel. Next, the bidirectional promoter fragment was digested with KpnI/BamHI and inserted between KpnI/BamHI digested SV40pA_Fluc pcDNA5/FRT/TO plasmid to generate SV40pA_Fluc_BI-16 pcDNA5/FRT/TO plasmid.

5) SV40pA was PCR amplified from pNL1.2 plasmid using primers (9/10). Next, the SV40pA fragment was digested with BclI/SphI restriction enzymes, and cloned between BclI/SphI sites in pcDNA5/FRT/TO plasmid to obtain SV40pA pcDNA5/FRT/TO plasmid.

6) *NlucP* sequence was PCR amplified from pNL1.2 plasmid using primers (11/12). Next, the *NlucP* fragment was digested with XhoI/ApaI restriction enzymes, and cloned between XhoI/ApaI sites in SV40pA_pcDNA5/FRT/TO plasmid to obtain NlucP_SV40pA pcDNA5/FRT/TO plasmid.

7) Oligo 13 and 14, as well as, oligo 15 and 16 were annealed to generate Linker 13/14 and ATGLinker 15/16, respectively. Linker 13/14 was inserted, in a correct orientation, between XhoI site in NlucP_SV40pA pcDNA5/FRT/TO plasmid to obtain Linker_NlucP_SV40pA pcDNA5/FRT/TO plasmid. ATGLinker 15/16 was inserted between BamHI/XhoI sites in NlucP_SV40pA pcDNA5/FRT/TO plasmid to obtain ATGLinker_NlucP_SV40pA pcDNA5/FRT/TO plasmid.

8) ATGLinker_NlucP_SV40pA pcDNA5/FRT/TO plasmid was digested with BamHI/SphI restriction enzymes to obtain ATGLinker_NlucP_SV40pA fragment. Next, the ATGLinker_NlucP_SV40pA insert was cloned between BamHI/SphI sites in SV40pA_Fluc_BI-16 pcDNA5/FRT/TO plasmid to obtain SV40pA_Fluc_BI-16_ATGLinker_NlucP_SV40pA pcDNA5/FRT/TO final plasmid.

9) Linker_NlucP_SV40pA pcDNA5/FRT/TO plasmid was digested with XhoI/SphI restriction enzymes to obtain Linker_NlucP_SV40pA fragment. Next, the Linker_NlucP_SV40pA insert was cloned between XhoI/SphI sites in SV40pA_Fluc_BI-16 pcDNA5/FRT/TO plasmid to obtain SV40pA_Fluc_BI-16_Linker_NlucP_SV40pA pcDNA5/FRT/TO plasmid.

10) Sequence of *HTT* exon 1 with 16 CAG or 98 CAG repeats were PCR amplified using (17/18) primers from previously obtained *HTT* cDNA clones (from Coriell Repositories cell line NM_002111). Next, both

the amplicons were digested with BamHI/XhoI restriction enzymes, and cloned between BamHI/XhoI restriction sites in SV40pA_Fluc_BI-16_Linker_NlucP_SV40pA pcDNA5/FRT/TO plasmid to obtain SV40pA_Fluc_BI-16_HTT16_Linker_NlucP_SV40pA pcDNA5/FRT/TO and SV40pA_Fluc_BI-16_HTT98_Linker_NlucP_SV40pA pcDNA5/FRT/TO final plasmids respectively.

Details of generation of Flp-In T-REx-293 cell lines

Flp-In T-REx-293 Host Cell line was grown in 21 cm^2 dishes to ~70% confluence. Next, the cells were co-transfected with a mixture containing pcDNA5/FRT/TO-based expression vector and pOG44 vector (1:9 ratio) using Lipofectamine 2000 transfection reagent (all from Invitrogen). The transfection mix, containing 9 µg of pOG44, 1 µg of the pcDNA5/FRT/TO-based expression vector, 1 ml of Opti-MEM and 20 µL of Lipofectamine 2000, was incubated for 30 min at RT and added directly to Flp-In T-REx-293 Host cells. At 48 h post-transfection, the medium was replaced with DMEM selection medium, additionally supplemented with 100 µg/ml hygromycin B and 5 µg/ml blasticidin S. The medium was replaced every 3 to 4 days to remove dead cells. After 2 to 3 weeks, individual hygromycin-resistant colonies were clonally selected using cloning cylinders (Sigma-Aldrich). The individual colonies were further expanded and stored in liquid nitrogen using the same medium supplemented with 10% DMSO. The expression of recombinant proteins was screened following induction of cells with doxycycline (Sigma-Aldrich) by using immunoblotting and bioluminescence analysis (Supplementary Figure S5).

Selection of single-cell clones for AGO2 knock-out and AGO2 mutant cell lines

GFP-positive cells were sorted into 96-well plate (one cell per well) using the BD FACSAria III (BD Biosciences) flow cytometer 48 h post-electroporation. The cells were cultured for about two weeks, after which genomic DNA was isolated with the QuickExtract DNA Extraction Solution (Lucigen Corporation) according to the manufacturer's instructions. Screening of clones harboring deletions (the pair of Cas9_sg1 and Cas9_sg2 caused deletion of 77 bp fragment leading to premature STOP codons) was carried out using GoTaq Polymerase (Promega) according to the manufacturer's instructions with PCR primers AGOe2_F and AGOe2_R. HDR-mediated clones were validated with an assay including PCR with AGOe14_F and AGOe14_R primers and digestion of PCR product using AflII and BseNI restriction enzymes (New England Biolabs). Positive clones (GAC to GCC codon change) should have restriction site for BseNI and lost restriction site for AflII enzyme. Selected clones were sequenced using primer AGOe14_F. Clones with confirmed frame shift or codon change were subjected to western blot analysis. Primer sequences are given in Supplementary Table S8.

Mycoplasma testing

All cell cultures used were routinely negatively verified for mycoplasma contamination using Veno GeM Classic Mycoplasma PCR detection Kit (Minerva Biolabs) according to manufacturer's instruction.

SUPPLEMENTARY TABLES

Supplementary Table 1. Publications describing activity of art-miRNAs

Targeted gene	**Model**	**Type of art-miRNA**	**Reference** (numbers according to main text references)
HTT	human fibroblasts	RNA duplexes	[27]
HTT	human fibroblasts	RNA duplexes	[28]
ATXN3	human fibroblasts	RNA duplexes	[33]
HTT	human fibroblasts, mouse striatal precursor cells	RNA duplexes	[34]
HTT	human fibroblasts, HD mouse model	chemically-modified single-stranded RNAs	[35]
HTT, ATXN3	human fibroblasts	chemically-modified RNA duplexes	[36]
HTT	human fibroblasts	self-duplexing RNAs	[37]
HTT, ATXN3	human fibroblasts	chemically-modified RNA duplexes	[38]
ATXN3	human fibroblasts	chemically-modified single-stranded RNAs	[39]
HTT	human fibroblasts, HD mouse model	chemically-modified single-stranded RNAs	[40]
ATN1	human fibroblasts	RNA duplexes and chemically-modified RNAs	[29]
HTT, *ATN1*, *ATXN3*	human fibroblasts, mouse striatal precursor cells	RNA duplexes, self-duplexing RNAs and its chemically modified versions	[30]
ATXN7	human fibroblasts	RNA duplexes, self-duplexing RNAs and its chemically modified versions	[31]
HTT	human fibroblasts	self-duplexing RNAs and its chemically modified versions	[32]

Supplementary Table 2. Sequences of RNA oligonucleotides. For art-miRNAs (A2, A4, G2 and G4) nucleotides which form mismatches in the interaction with targeted CAG repeat tract are in bold.

Name	**Sense (5'-3')**	**Antisense (5'-3')**
A2	GCUGCUGC**A**GCUGCUGCUGCU	GCUGCUGC**A**GCUGCUGCUGCU
siHTT*	GCCUUCGAGUCCCUCAAGUCC	ACUUGAGGGACUCGAAGGCCU
siRluc	AUCUGAAGAAGGAGAAAAATT	AUCUGAAGAAGGAGAAAAATT
siFluc	CGUACGCGGAAUACUUCGAUU	UCGAAGUAUUCCGCGUACGUU
A4	GCUGCUGC**A**GCUGC**A**GCUGCU	GCUGCUGC**A**GCUGC**A**GCUGCU
G2	GCUGCUGC**G**GCUGCUGCUGCU	GCUGCUGC**G**GCUGCUGCUGCU
G4	GCUGCUGC**G**GCUGC**G**GCUGCU	GCUGCUGC**G**GCUGC**G**GCUGCU

* from published study, [4].

Supplementary Table 3. Antibodies

Name (host)	Dilution	Company (catalogue number)
Western blot		
anti-huntingtin (mouse)	1:1000	Millipore (MAB2166)
anti-atrophin-1 (rabbit)	1:1000	Bethyl Laboratories (A300-753A)
anti-ataxin-3 (mouse)	1:1000	Millipore (MAB5360)
anti-GAPDH (mouse)	1:10000	Millipore (MAB374)
anti-vinculin (rabbit)	1:1000	Cell Signaling Technology (4650)
anti-plectin (rabbit)	1:1000	Abcam (ab83497)
anti-firefly luciferase (rabbit)	1:1000	Thermo Fisher Scientific (PA5-32208)
anti-rabbit HRP-conjugate	1:2000	Jackson ImmunoResearch (711-035-152)
anti-mouse HRP-conjugate	1:2000	Jackson ImmunoResearch (715-035-150) or Sigma-Aldrich (A9917)
Immunocytochemistry		
anti-Pax6 (rabbit)	1:100	Cell Signaling Technology (60433)
anti-Sox1 (rabbit)	1:200	Cell Signaling Technology (4154)
anti-Sox2 (rabbit)	1:200	Cell Signaling Technology (3579)
anti-nestin (mouse)	1:500	Stem Cell Technology (60091)
anti-rabbit Alexa Fluor 488	1:1000	Jackson ImmunoResearch (711-546-152)
anti-mouse Alexa Fluor 594	1:1000	Jackson ImmunoResearch (715-586-151)

Supplementary Table 4. Sequences of DNA oligonucleotides used as PCR primers in RT-qPCR, ddPCR and poly(A) tail length assay.

Name	Forward (5'-3')	Reverse (5'-3')
GAPDH	GAAGGTGAAGGTCGGAGTC	GAAGATGGTGATGGGATTTC
Fluc	AAGAAGTGCTCGTCCTCG	TGATCATGAGCGGCTACG
NlucP	AAGGTGATCCTGCACTATGGC	TCTTTTTGCCGTCGAACACG
ATXN1	CAGCCCTGTCCAAACACAAA	GCAACGACCTGAAGATCGAC
ATXN3	AGTTCAGGAGCACTTGGGAG	CAAAGTGGACCCTATGCTGT
ATXN7	AGGTGTTCTTAGCGCATCCT	AGTGTGCCATCCATTTTCGG
ATN1	TGCTATCCATGCAGCCTCTG	AGCAAAGAGCTGGTGACGAA
HTT	CGACAGCGAGTCAGTGAATG	ACCACTCTGGCTTCACAAGG
PAX6	TGCTCCGGCATGAAATATACTA	GTCTCCAAATGTGCAGCAAC
SOX1	ACCAGGCCATGGATGAAG	CTTAATTGCTGGGGAATTGG
NES	GCGGGCTACTGAAAAGTTCC	CAGGAGGGTCCTGTACGTG
HTT-NlucP - poly(A) length assay	GCTGTTCCGAGTAACCATCAA CG	GCAATAGCATCACAAATTTCACAAA TAAAGC

Supplementary Table 5. Sequences of oligonucleotides for *ATXN3* smFISH (sequence 5'- 3')

1	TCGTGGAAGATGGACTCCAT
2	ACAAAGTGAGCCTTCTTGTT
3	AGTTATTCAGGCAATGTTGA
4	GGCTAAAATATTCTCCTTGC
5	CAGCTGATGTGCAATTGAGG
6	ATCTTCACTAGTAACTCCTC
7	GCTGCTGTAAAAACGTGCGA
8	AACCACTGTCATCCATATTT
9	AGGCATTGCTTATAACCTGA
10	GGATTAGTTCTAAACCCCAA
11	CTGATACTCTGGACTGTTGA
12	TTATAGGATCGATCCTGAGC
13	CTGTAAACCAGTGTTCCTTA
14	TTAATTCTGGACCCGTCAAG
15	GAGCCAAGAAAAGTGCAAGA
16	AGAATAACCTTCCTGTTGTA
17	GCAGATCACCCTTAACAACA
18	AGTTGGTCAGCTTCGCAATC
19	TTGGACCCTAATCATCTGCA
20	TAAGTTTTGGTCGATGCATC
21	AGTTGTGCTAATTCTTCTCC
22	GTTTTATGGACTCTTTGCTC
23	CTTCTAACATTCGTTCCAGG
24	TAACATTCCTGAGCCATCAT
25	TGCAAATCCTCCTCATCTTC
26	TGTCAATTTCTTGGCGACTT
27	AGATCTGCTTCCTCATCTTC
28	TACTTAGCTGAATAGCCCTG
29	ATGTTTCTGGAACTACCTTG
30	TACCTGATGTCTGTGTCATA
31	TCCGAAGCTCTTCTGAAGTA
32	GATGTGAACTCTGTCCTGAT
33	CAAGTGCTCCTGAACTGGTG
34	ATAGCATCACCTAGATCACT
35	CTGAAGCATGTCTTCTTCAC
36	CTAAAGACATGGTCACAGCT

Supplementary Table 6. Sequences of DNA oligonucleotides used for cloning for generation of constructs used for transient expression (relative to Figure 2)

Name	Forward/sense (5'-3')	Reverse/antisense (5'-3')
Mutagenesis		
5'ORF EcoRI	GGCAATCCGGTACTGTTGGTAA A**GAATTC**GCCACCATGGAA	TTCCATGGTGGC**GAATTC**TTTACCAA CAGTACCGGATTGCC
5'ORF EcoR*I*/NdeI	GGTAAA**GAATTC**GCCACCATG**C ATATG**GAAGATGCCAAAAACATT AAG	CTTAATGTTTTTGGCATCTTC**CATAT G**CATGGTGGC**GAATTC**TTTACC
DelA in STOP codon TAA(TA^T)	GCGGCAAGATCGCCGTG**TAT**TC TAGTTGTTTAAAC	GTTTAAACAACTAGA**ATA**CACGGCGA TCTTGCCGC
Plasmid generation		
ATXN3 3'UTR/3'ORF XbaI/SalI	CGCCATCTACGGGTA**TCTAGA** GAAGAGCTTCGGAAGAGAC	GATCTGACTAATGCT**GTCGAC**TTATC CTGAACTGGTGGCTGGC
ATXN3 5'ORF EcoRI/NdeI	GAAATCTACGGGTATC**GAATTC** GCCACCATGGCGGAAGAGCTTC GGAAGAGAC	AGGATCTACGGGTAA**CATATG**TCCT GAACTGGTGGCTGGC
HTT 3'UTR/3'ORF XbaI/SalI	GATCATCTACCGTTA**TCTAGA**AT GGCGACCCTGGAAAAGCTG	GATCTGACTAATGCTTT**GTCGAC**TTA CAGCGGCGGCGGCTGAGG
HTT 5'ORF EcoRI/NdeI	GAAATCTACGGGTATC**GAATTC** GCCACCATGGCGACCCTGGAAA AGCTG	AGATCTGACTAATGCTTT**CATATG**CA GCGGCGGCGGCTGAGG
Synthetic inserts		
CAG17 3'UTR/3'ORF (Xba*I*/Sal*I*)	**CTAGA**ATG$(CAG)_{17}$TAA**G**	**TCGAC**TTA$(CTG)_{17}$CAT**T**
CAG17 5'ORF (EcoR*I*/Nde*I*)	**AATTC**GCCACCATGGAG$(CAG)_{17}$ **CA**	**TATG**$(CTG)_{17}$C TCCATGGTGGC**G**

Supplementary Table 7. Sequences of DNA oligonucleotides used for cloning for generation of constructs used for stable expression (relative to Figure 4)

ID	Oligo Name	Sequence (5'-3')
1	Fluc_KpnI_F	GCGGTACCGCCACCATGGAAGATGCCAAAAACATTAAG
2	Fluc_BamHI_R	CGGGATCCTCATTACACGGCGATCTTGCCGCCCTTCTTG
3	SV40pA_BglII_F	GAAGATCTCAGACATGATAAGATACATTGATG
4	SV40pA_BamHI_R	ATGTGCGGCCGCGGATCCTACCACATTTGTAGAGGTTTTACTTGC
5	BI16_F1_XbaI_F	GCTCTAGAGTACATCAAGTGTATCAT
6	BI16_F1_KpnI_R	GGCGGTACCGACGATCTCTATCACTG
7	BI16_F2_XbaI_F	GCTCTAGAAAATGTCGTAACAACT
8	BI16_F2_BamHI_R	GGCGGATCCGACGATCTCTATCACTG
9	SV40pA_BclI_F	GCGCTGATCACAGACATGATAAGATACATTGATG
10	SV40pA_SphI_R	CCGTGCATGCTACCACATTTGTAGAGGTTTTACTTGC
11	NlucP_XhoI_F	CCGCTCGAGATGGTCTTCACACTCGAAGATTTCG
12	NlucP_ApaI_R	CTGGGCCCTCATTAGACGTTGATGCGAGCTGAAGCAC
13	Linker 13/14 ss	(Phosp)TCGAGCATCATCACCACCATCATGAGAACCTGTACTTTCAGAGCGATTACAAGGATGACGACGATAAGT
14	Linker 13/14 as	(Phosp)TCGAACTTATCGTCGTCATCCTTGTAATCGCTCTGAAAGTACAGGTTCTCATGATGGTGGTGATGATGC
15	ATGLinker 15/16 ss	(Phosp)GATCCGCCACCATGGATTACAAGGATGACGACGATAAGC
16	ATGLinker 15/16 as	(Phosp)TCGAGCTTATCGTCGTCATCCTTGTAATCCATGGTGGCG
17	HTT_BamHI_F	CGGGATCCGCCACCATGGCGACCCTGGAAAAGCTGATG
18	HTT_XhoI_R	GCATCTCGAGTGGTCGGTGCAGCGGCTCCTCAGC
19	AGO2_HindIII_F	CCCAAGCTTGCCGCCATGGACTACAAGGACG
20	AGO2_BamHI_R	GGCGGATCCTTATCAAGCAAAGTACATGGTGCGCAG

Supplementary Table 8. Sequences of DNA oligonucleotides used for sgRNA constructs and validation of genomic DNA editing

ID	Sequence 5'- 3'	Description
sgRNA1s	CACCGAGGCTTGAAGGCATATCCT	oligo for Cas9_sg1 plasmid construction
sgRNA1a	AAACAGGATATGCCTTCAAGCCTC	oligo for Cas9_sg1 plasmid construction
sgRNA2s	CACCGACAGGCCAATTTCTTCGAAA	oligo for Cas9_sg2 plasmid construction
sgRNA2a	AAACTTTCGAAGAAATTGGCCTGTC	oligo for Cas9_sg2 plasmid construction
sgRNA3s	CACCGTCTGCTCCCAGAAAGATGA	oligo for Cas9_sg3 plasmid construction
sgRNA3a	AAACTCATCTTTCTGGGAGCAGAC	oligo for Cas9_sg3 plasmid construction
AGOe2_F	AATGCAATATTGGCCTGGAC	forward primer for exon2 of *Ago2* gene
AGOe2_R	CGCAGACCACTTACACAGGT	reverse primer for exon2 of *Ago2* gene
AGOe14_F	CACCTCCAAAAGGTGATGGT	forward primer for exon2 of *Ago2* gene
AGOe14_R	AGCTTAGTGAGACCCCATGC	reverse primer for exon2 of *Ago2* gene
U6-Fwd	GAGGGCCTATTTCCCATGATTCC	sequencing primer for sgRNA validation
ssODN	GACGCCACTCCTCTCCGCAGGCCGCCGGTGTTCCAGCAGCCAGTCATCTTTCTGGGAGCAGCCGTCACTCACCCCCCCGCCGGGGATGGGAAGAAGCCCTCCATTGCCGCCGTGAGTGTCAGC	donor template for HDR

SUPPLEMENTARY REFERENCES

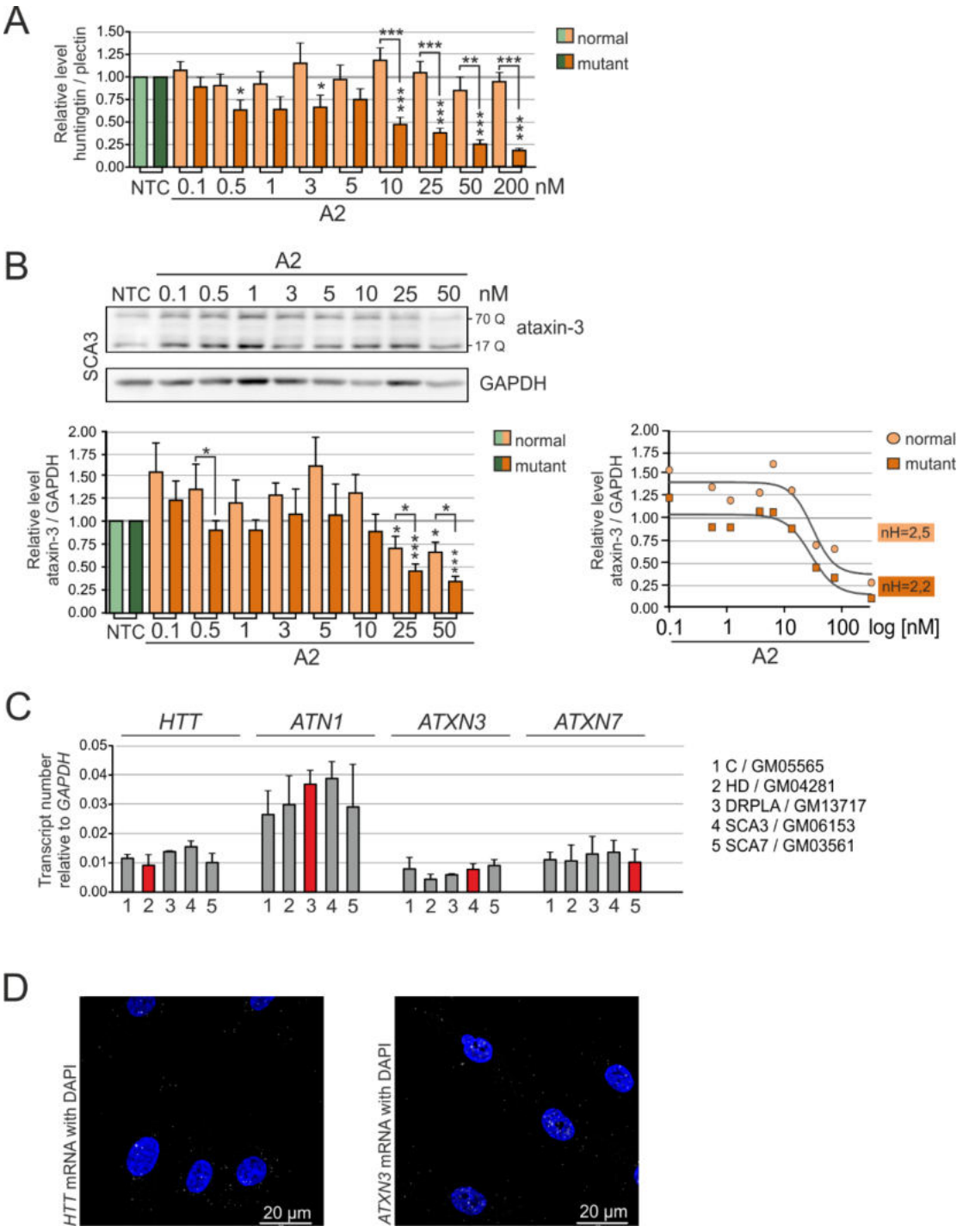
1. Chang T-H, Huang H-Y, Hsu JB-K, et al (2013) An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs. BMC Bioinformatics 14 Suppl 2:S4.

2. Gu S, Jin L, Zhang F, et al (2009) The biological basis for microRNA target restriction to the 3' untranslated region in mammalian mRNAs. Nat Struct Mol Biol 16:144–150.

3. Sammarco MC, Grabczyk E (2005) A series of bidirectional tetracycline-inducible promoters provides coordinated protein expression. Anal Biochem 346:210–6.

4. Wang YL, Liu W, Wada E, et al (2005) Clinico-pathological rescue of a model mouse of Huntington's disease by siRNA. Neurosci Res 53:241–249

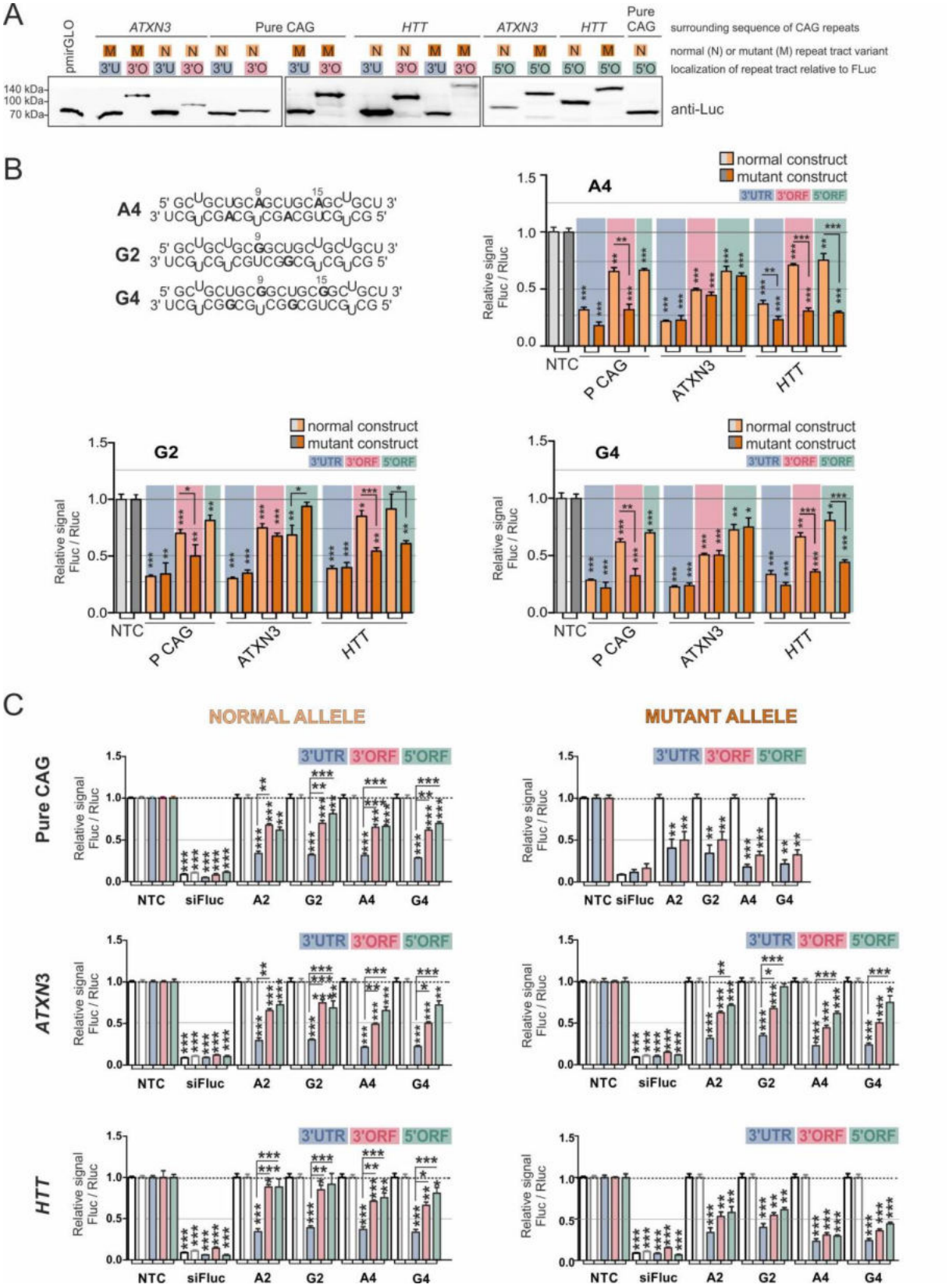
SUPPLEMENTARY FIGURES



Supplementary Figure 1. Additional data for Figure 1

(A) Detailed results corresponding to western blot analysis of huntingtin levels in HD-patient derived fibroblasts presented in Fig. 1C. NTC – cells treated with non-targeting siRNA (BlockIT siRNA). n=3 **(B)** Western blot analysis of ataxin-3 levels in SCA3 fibroblasts lysed 72 h after transfection with indicated concentration of A2. Results are presented as dose-response curves which were used for indicated Hill's coefficient calculation. NTC – cells treated with non-targeting siRNA (BlockIT siRNA). For (A) and

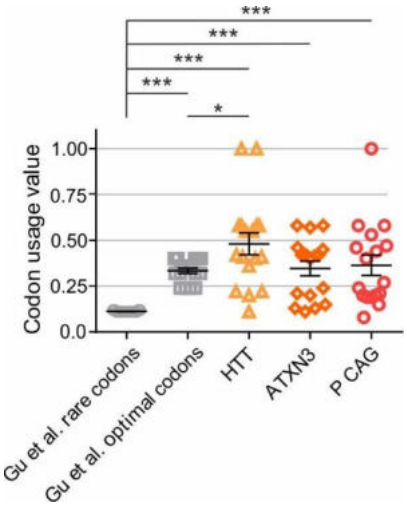
(B) data were analyzed using one-way ANOVA (with Bonferroni multiple comparisons test) n=3 **(C)** ddPCR-based quantification of indicated transcripts number in indicated fibroblast cell lines (disease and repository number are given). Two isolations of total RNA were performed from independent cultures of fibroblasts to obtain two sets of cDNA for this analysis. Red bars represent result obtained in cell line with mutant transcript. **(D)** Representative smFISH images for *HTT* (left panel) and *ATXN3* (right panel) mRNAs in healthy person-derived fibroblasts. DAPI was used for nuclear staining.



Supplementary Figure 2. Additional data for Figure 2

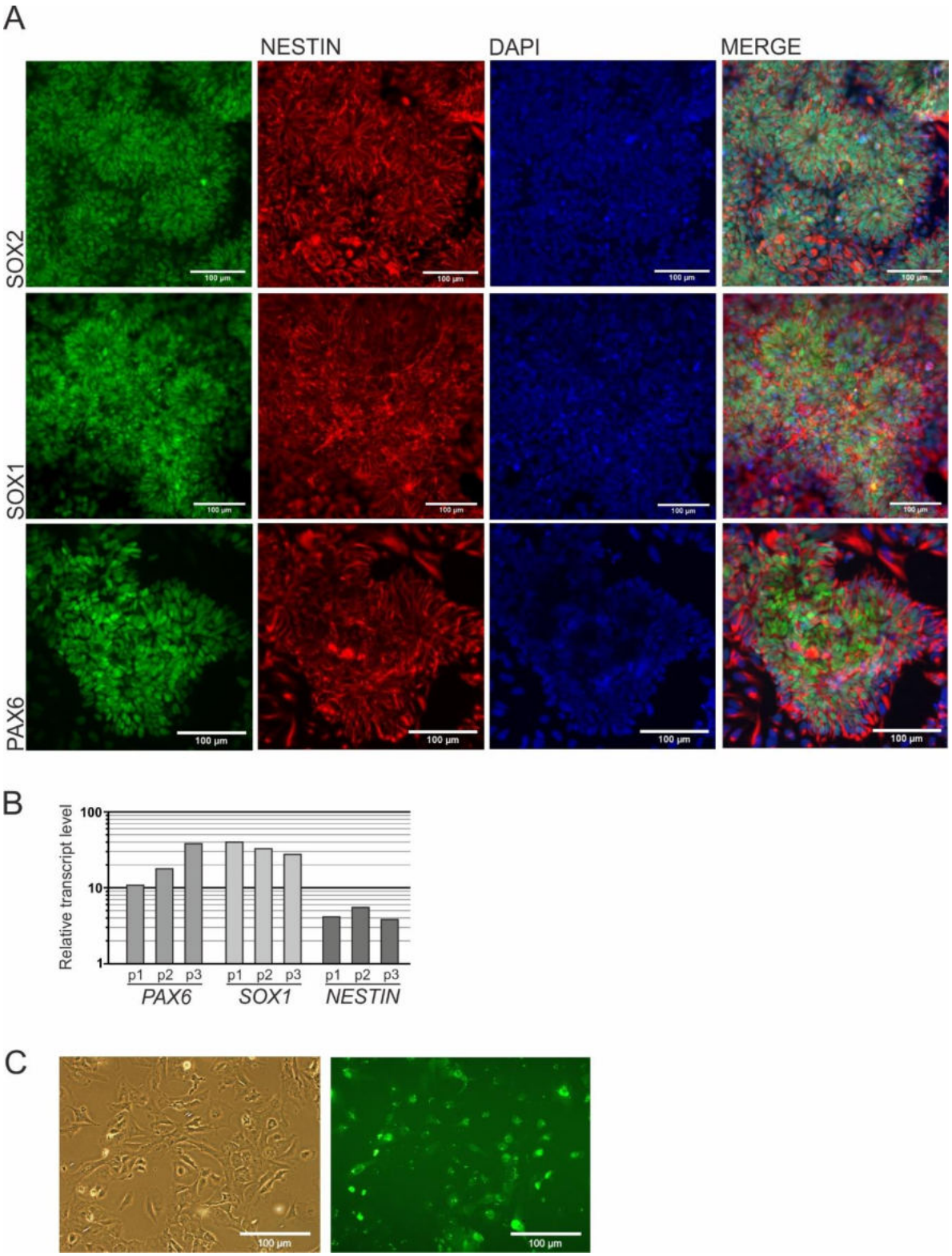
(A) Western blot for Fluc, expressed in HEK 293T cells transfected with designed pmirGLO-based constructs. Fusion proteins include addition of normal or mutant CAG repeat tract length, various

location of repeat tract (3'U - 3'UTR, 3'O - 3' site of ORF, 5'O - 5' site of ORF), as well as different repeat tract surrounding sequence. **(B)** Sequences of A4, G2 and G4 art-miRNAs and predicted base-pairing of two strands within a duplexes. Luciferase assay performed 24 h after cotransfection of HEK 293T cells with 50 nM A4, G2 or G4 and 100 ng of indicated plasmids. C – light grey: cells treated with pmirGLO plasmid and non-targeting siRNA (NTC), dark grey: “5’ORF”-modified pmirGLO plasmid treated with non-targeting siRNA (signal normalization details are given in Materials and Methods). Data was analyzed using one-way ANOVA (with Bonferroni multiple comparisons test). n=3 **(C)** Different representation of data shown in Fig. 2C and panel (B) of this figure. Graphs present activity of a set of art-miRNAs (A2, G2, A4 and G4), together with results for siRNA targeting *Fluc*, for selected construct: with normal repeat tract (left panels) or mutant one (right panels) and with various repeat tract surrounding sequence (Pure CAG - no additional sequence added or from *HTT* or *ATXN3* genes). n=3



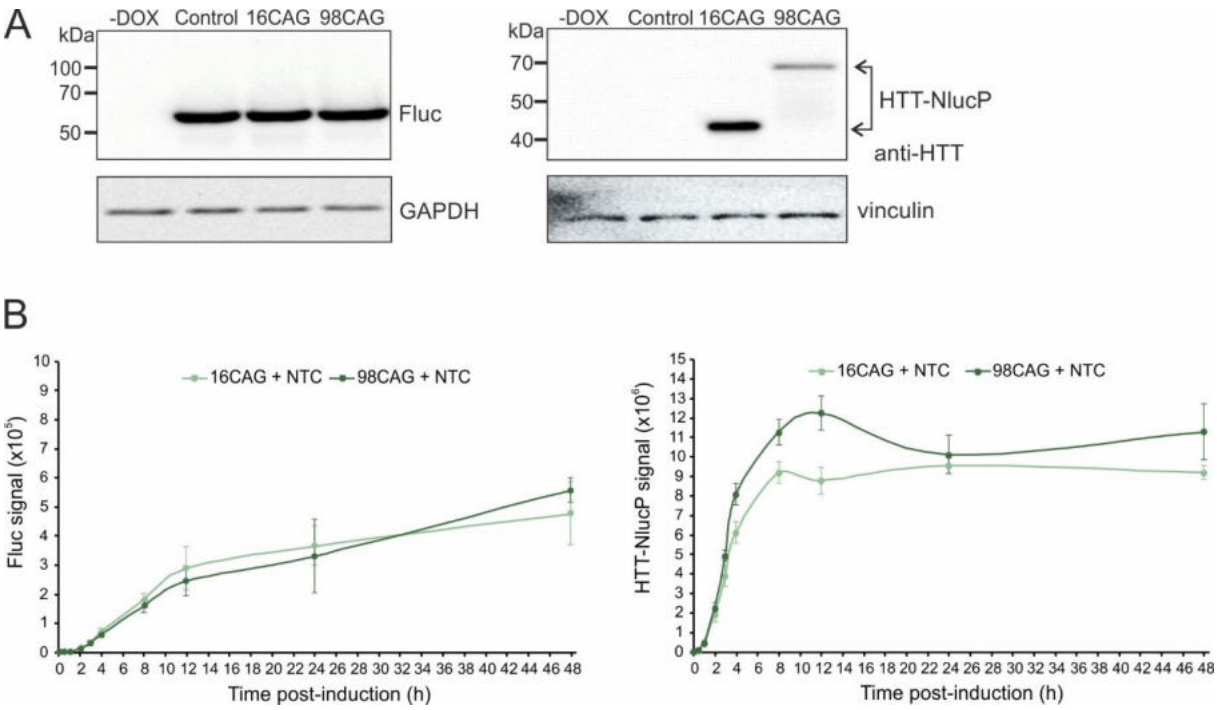
Supplementary Figure 3. Codon usage bioinformatic analysis

Codon usage of 17 codons upstream to CAG repeat tract in *HTT, ATXN3* and in 3'ORF *Fluc* (upstream to MCS and STOP codon of pmirGLO). As controls we used codon described as rare or optimal in Gu et al., 2009. Data was analyzed using one-way ANOVA (Bonferroni multiple comparisons test).



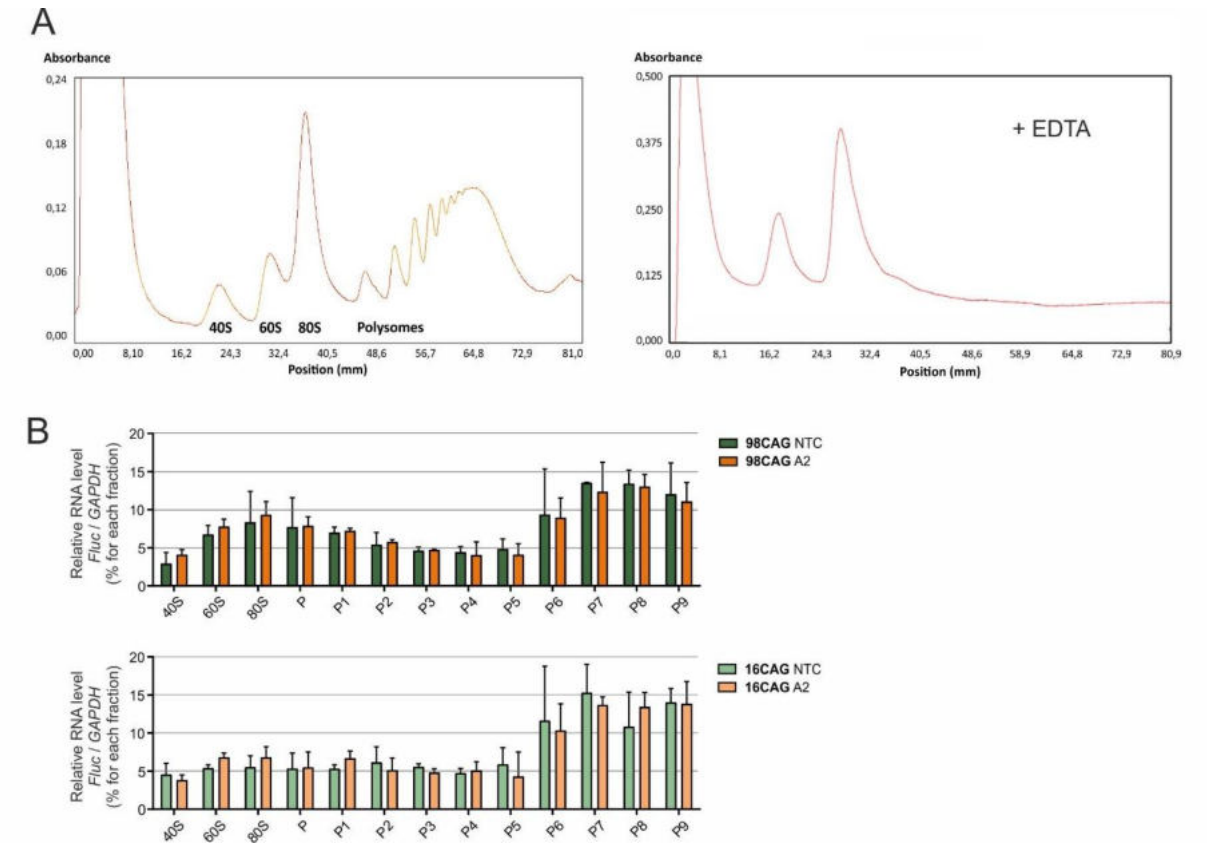
Supplementary Figure 4. Additional data for Figure 3

(A) Immunofluorescence staining of HD NPs for selected markers for neural stem cells. **(B)** RT-qPCR results of expression level for selected neural stem cells markers: *PAX6*, *SOX1* and *NESTIN* in HD NPs, analyzed at passage 1, 2 and 3 (indicated as p1, p2 and p3). Results are presented as mean relative to expression level in parental HD iPSCs, set at 1, and were normalized to GAPDH expression level. **(C)** Exemplary images of HD NPs 24 h after transfection of 100 nM BlockIT with siPORT Amine.



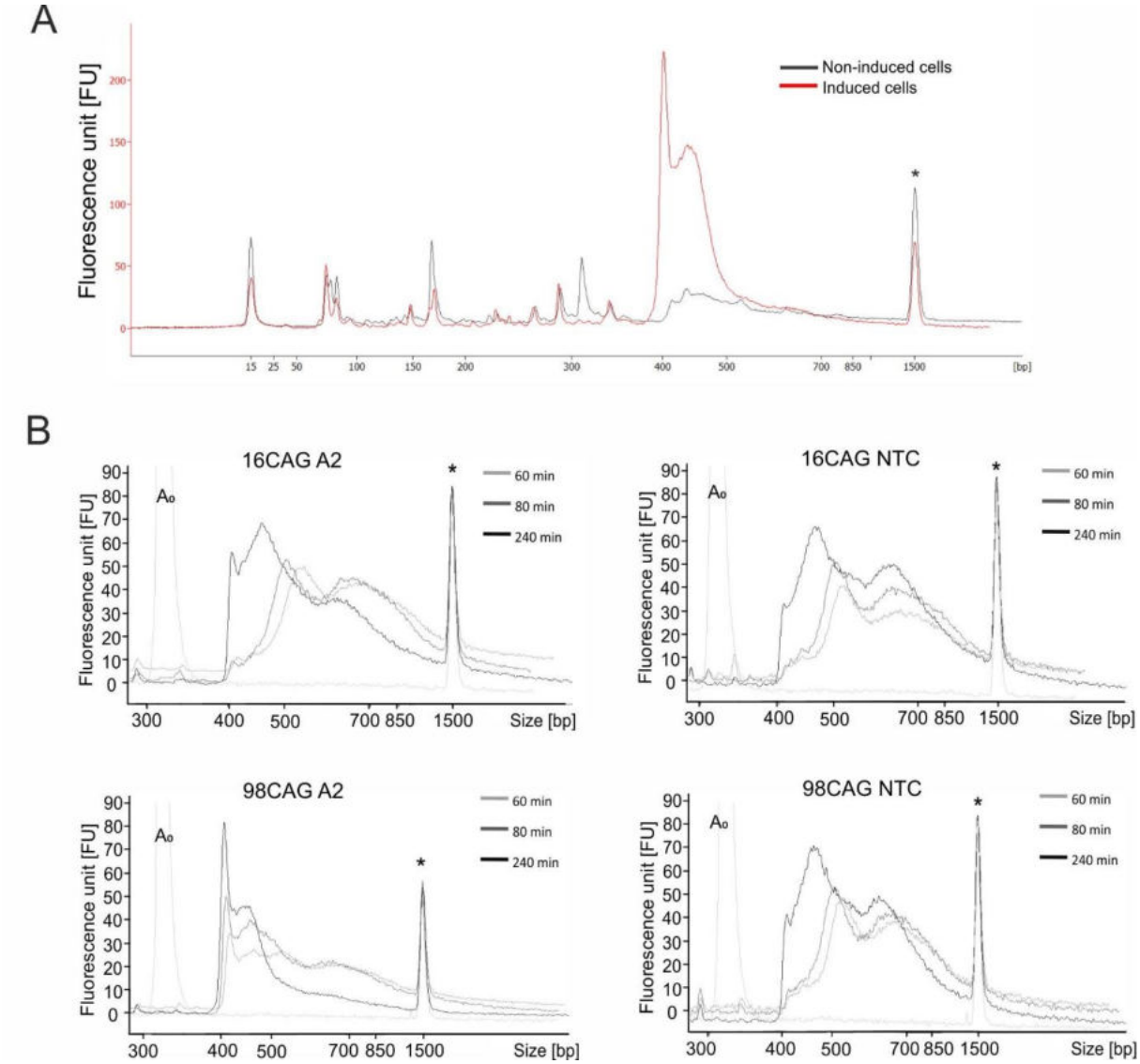
Supplementary Figure 5. Additional data for Figure 4: Flp-In T-REx-293 cell lines

(A) Western blot for Fluc (left panel) and HTT-NlucP fusion protein in generated 16CAG and 98CAG Flp-In T-REx-293 cell lines, as well as control cell line (lacking CAG repeat insert). Control, 16CAG and 98CAG lysates were prepared 24 h after induction with doxycycline. GAPDH or vinculin were used as protein loading controls. **(B)** Time course of Fluc (left panel) and HTT-NlucP (right panel) signal after induction. 16CAG and 98CAG cell lines were transfected with control siRNA (NTC, siRluc) and 12 h later expression was induced with doxycycline. Cells were harvested at 30 min, 1, 2, 3, 4, 8, 12, 24 and 48 h points, counted and the same number of cells were lysed and processed for Nano-Glo Dual-Luciferase Reporter Assay. Luciferase activity is given in arbitrary units. n=3



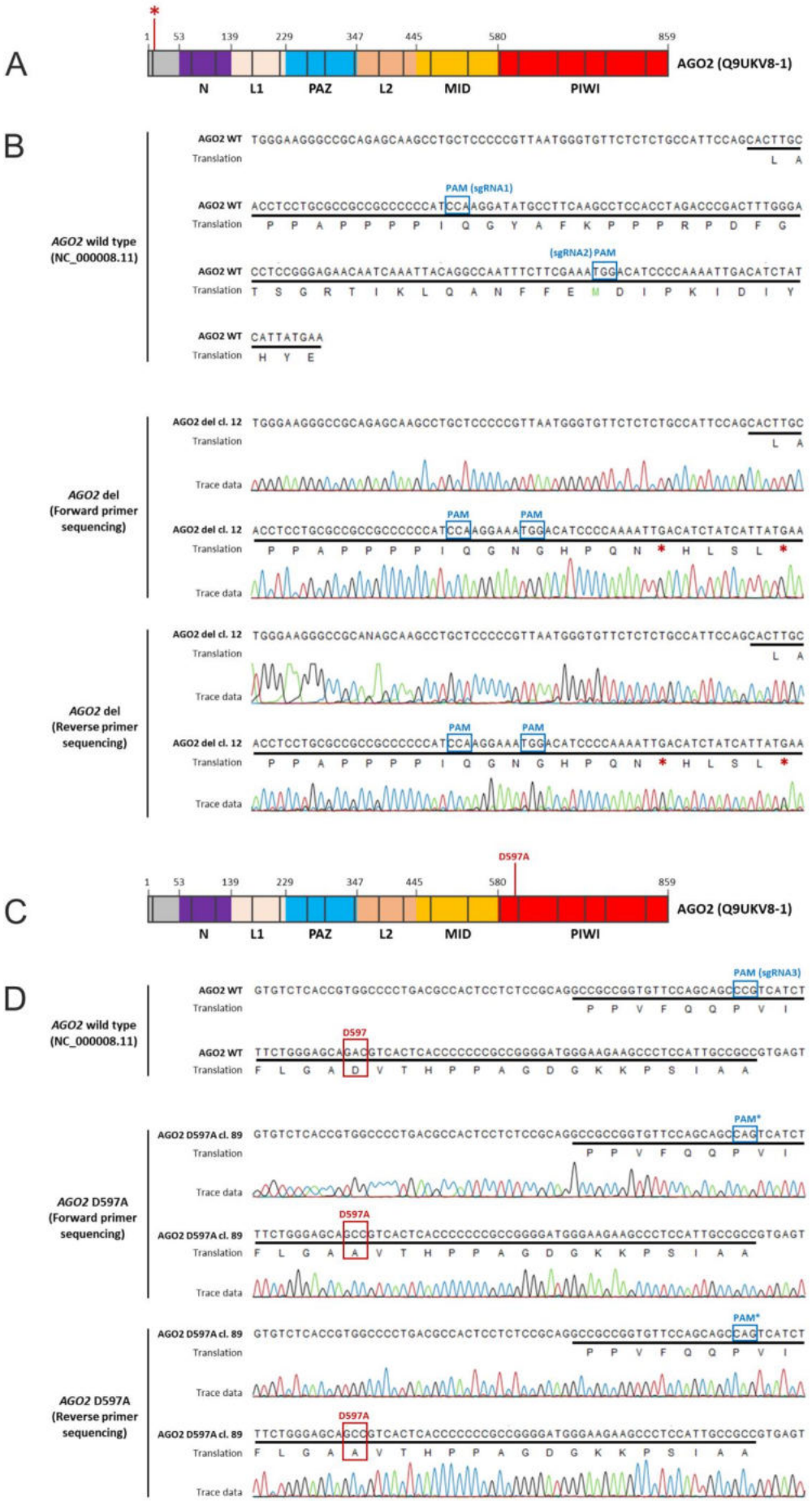
Supplementary Figure 6. Additional data for Figure 4: polysome profiling

(A) Left panel: exemplary graph of A260 profile obtained in sucrose gradient separation of Flp-In T-REx-293 cell line lysate, right panel: analogous results obtained after addition of 20 nM of EDTA to lysis buffer and sucrose solutions. **(B)** RT-qPCR results of control *Fluc* expression level after transfection of 98CAG (upper panels) or 16 CAG (lower panels) cell line with 100 nM A2 or control siRNA (NTC, siRLuc), at 3 h time point after induction, in indicated fractions of ribosomal subunits (40S and 60S), 80S monosome and polysomes (P-P9). Data were normalized to *GAPDH* expression in each fraction and are presented as % share of *HTT-NlucP* expression where 100 % is the sum of obtained values for all fractions. Data analysis using two-way ANOVA showed lack of statistically significant results in comparison of *Fluc* expression after treatment. n=3



Supplementary Figure 7. Additional data for Figure 4: poly(A) tail-length assay

(A) Example of raw data for resolution of products obtained from poly(A) tail length assay for 98CAG cell line: non-induced cells and cells 4 h after induction and treatment with A2. **(B)** Different graphs containing data from the Figure 4F for clear presentation of time-course changes in representation of poly(A) tails for specific cell line and after specific siRNA treatment: analysis of poly(A) tails lengths of *HTT-NlucP* transcript in 98CAG cells at indicated time points (60, 80, 120 min) after transfection with 100 nM A2 or control siRNA (NTC, siRLuc). Estimated poly(A) tails lengths are indicated. Experiment was repeated and similar results were obtained. A_0 - peak obtained with reporter-specific primers amplifying a region upstream of the polyadenylation site. * - an internal standard peak (1500 bp Upper Marker).


A
1
53
139
229
347
445
580
859
AGO2 (Q9UKV8-1)
N
L1
PAZ
L2
MID
PIWI
B
AGO2 WT
Translation
PAM (sgRNA1)
(sgRNA2) PAM
AGO2 wild type
(NC_000008.11)
AGO2 del cl. 12
Trace data
PAM
AGO2 del
(Forward primer
sequencing)
AGO2 del
(Reverse primer
sequencing)
C
D597A
D
PAM (sgRNA3)
D597
AGO2 D597A cl. 89
PAM*
AGO2 D597A
(Forward primer
sequencing)
AGO2 D597A
(Reverse primer
sequencing)

Supplementary Figure 8. Additional data for Figure 5. AGO2 knockout and mutant cell lines

(A) and **(C)** Schematic architecture of AGO2 protein (Q9UKV8-1) with characteristic protein domains (N, L1, PAZ, L2, MID, PIWI). The positions of CRISPR/Cas9-mediated knockout **(A)** and D597A mutation **(C)** are indicated by red lines. **(B)** DNA sequencing confirmation of correct homozygous AGO2 knockout in 98CAG Flp-In T-REX-293 cell line (AGO2del). Wild type *AGO2* reference sequence (NC_000008.11), as well as, chromatograms from forward and reverse primer DNA sequencing obtained from genomic DNA isolated from AGO2del clone 12 are shown. Underlined DNA sequence corresponds to exon 2. PAM sequences for sgRNA1 and sgRNA2 are indicated by blue frames. As a consequence of CRISPR/Cas-9 mediated knockout a deletion of 77 bp fragment is observed leading to the generation of premature termination codons indicated by asterisks. **(D)** DNA sequencing confirmation of correct homozygous AGO2(D597A) mutant 98CAG Flp-In T-REX-293 cell line (AGO2mut). Underlined DNA sequence corresponds to exon 14. Wild type *AGO2* reference sequence (NC_000008.11), as well as, chromatograms from forward and reverse primer DNA sequencing obtained from genomic DNA isolated from AGO2mut clone 89 are shown. PAM sequence for sgRNA3 is indicated by blue frame. As a consequence of CRISPR/Cas-9 mediated HDR-depended insertion D597A mutation (GAC to GCC) is observed, indicated by red frame. To prevent CRISPR/Cas-9 mediated re-cleavage a silent mutation in the PAM sequence was introduced (PAM*).

cells

MDPI

Review

Implications of Poly(A) Tail Processing in Repeat Expansion Diseases

Paweł Joachimiak, Adam Ciesiołka, Grzegorz Figura and Agnieszka Fiszer *

Department of Medical Biotechnology, Institute of Bioorganic Chemistry Polish Academy of Sciences, 61-704 Poznań, Poland; pjoachimiak@ibch.poznan.pl (P.J.); a.k.ciesiolka@gmail.com (A.C.); grzegorz.figura5@gmail.com (G.F.)

* Correspondence: agnieszka.fiszer@ibch.poznan.pl

Abstract: Repeat expansion diseases are a group of more than 40 disorders that affect mainly the nervous and/or muscular system and include myotonic dystrophies, Huntington's disease, and fragile X syndrome. The mutation-driven expanded repeat tract occurs in specific genes and is composed of tri- to dodeca-nucleotide-long units. Mutant mRNA is a pathogenic factor or important contributor to the disease and has great potential as a therapeutic target. Although repeat expansion diseases are quite well known, there are limited studies concerning polyadenylation events for implicated transcripts that could have profound effects on transcript stability, localization, and translation efficiency. In this review, we briefly present polyadenylation and alternative polyadenylation (APA) mechanisms and discuss their role in the pathogenesis of selected diseases. We also discuss several methods for poly(A) tail measurement (both transcript-specific and transcriptome-wide analyses) and APA site identification—the further development and use of which may contribute to a better understanding of the correlation between APA events and repeat expansion diseases. Finally, we point out some future perspectives on the research into repeat expansion diseases, as well as APA studies.

Keywords: alternative polyadenylation; repeat expansion diseases; polyglutamine diseases; Huntington's disease; poly(A) tail

Citation: Joachimiak, P.; Ciesiołka, A.; Figura, G.; Fiszer, A. Implications of Poly(A) Tail Processing in Repeat Expansion Diseases. *Cells* **2022**, *11*, 677. https://doi.org/10.3390/cells11040677

Academic Editor: Adrian Israelson

Received: 17 January 2022
Accepted: 13 February 2022
Published: 15 February 2022

1. Introduction

The main role of mRNAs is to transfer genetic information from DNA into proteins; however, mRNA also functions to provide an additional level of gene expression regulation in cells. Newly generated transcripts-before becoming translation templates-undergo modifications and processing, including splicing and the site selection of the transcript's 3′ end by the generation of a poly(A) tail [1]. As a result of the latter, a single mature mRNA has a defined 3′-UTR region and a specific poly(A) tail length. Nevertheless, the overall pool of generated mRNAs contains molecules with different 3′ ends. These features differ for any given mRNA-especially in the developmental stages and across various cell types-which affects its stability and overall transcript functioning.

There is a great need for the precise, molecular characterization of transcript variants to understand cell-type-specific processes and their disruption in cases of gene mutations. This includes aspects of poly(A) tail processing, which is especially crucial in neuronal cells, as features of the 3′ end of an mRNA play an important role in regulating the dynamics of RNA metabolism and translation [2].

In this review, we discuss poly(A) tail processing in a context of a group of neurodegenerative and neuromuscular diseases caused by genes mutations consisting of the expansion of a sequence of repeats, e.g., CAG or CUG. A better understanding of a poly(A) tail processing in the context of neuromuscular pathology may provide valuable insight into molecular pathomechanisms triggered by transcript misprocessing and help design RNA-targeted therapies.

We briefly present some general information regarding the processes associated with polyadenylation and the methods used for their analysis. We also discuss previously collected information concerning poly(A) tail processing in mRNAs implicated in repeat expansion diseases, as well as the identified disruptions in this process. This includes the analysis of alternative polyadenylation sites for a group of transcripts containing expanded CAG repeat tracts implicated in polyglutamine (polyQ) diseases.

2. Polyadenylation in Animal Cells

2.1. Polyadenylation Process

During polyadenylation, a polyadenosine sequence-namely, a poly(A) tail-is added to the 3′ end of a transcript. Together with the removal of introns and the addition of a 5′ cap, polyadenylation constitutes a major step in pre-mRNA maturation [1,3,4]. The polyadenylation process can be divided into two major steps: first, newly transcribed pre-mRNA is cleaved and its 3′ end is generated; then, a specific enzyme-poly(A) polymerase (PAP)-generates the poly(A) tail independently from the template, starting from the cleavage site [5–7]. These two steps depend on the interplay between various *cis*- and *trans*-acting factors (Figure 1). *Cis* elements, which are typically located in 3′-UTRs, determine the position of cleavage and polyadenylation site in mRNA [8]. The most crucial *cis*-element is an RNA sequence motif called the polyadenylation signal (PAS). In animals, the canonical PAS is the hexanucleotide sequence AAUAAA [9–11]. The cleavage site is located around 10 to 30 nucleotides downstream from the PAS, between the PAS and a core downstream sequence element (DSE)-a U-/GU-rich sequence, which is itself located 14–70 nucleotides downstream of the cleavage site (Figure 1) [12]. The cleavage itself occurs just before an adenosine residue, mostly after cytosine. Another *cis*-element that is involved in the polyadenylation process is a U-rich/UGUA upstream sequence element (USE) located upstream of PAS. All three *cis*-elements are recognized by their respective *trans*-acting factors, which interact with each other through the carboxyl-terminal domain (CTD) of RNA polymerase II (RNAPII) [5]. PAS recruits cleavage and polyadenylation specificity factor (CPSF), DSE is recognized by cleavage stimulation factor (CstF), and USE recruits cleavage factor I_m (CFI_m). Cooperation between those factors leads to cleavage. The subsequent cooperation between nuclear poly(A)-binding protein (PABPN1) and PAP allows for the generation of the poly(A) tail at the cleavage site: PABPN1 acts as a type of 'ruler', which is important for the synthesis of an appropriately sized poly(A) tail, while PAP performs a non-templated addition of adenosine residues [13].

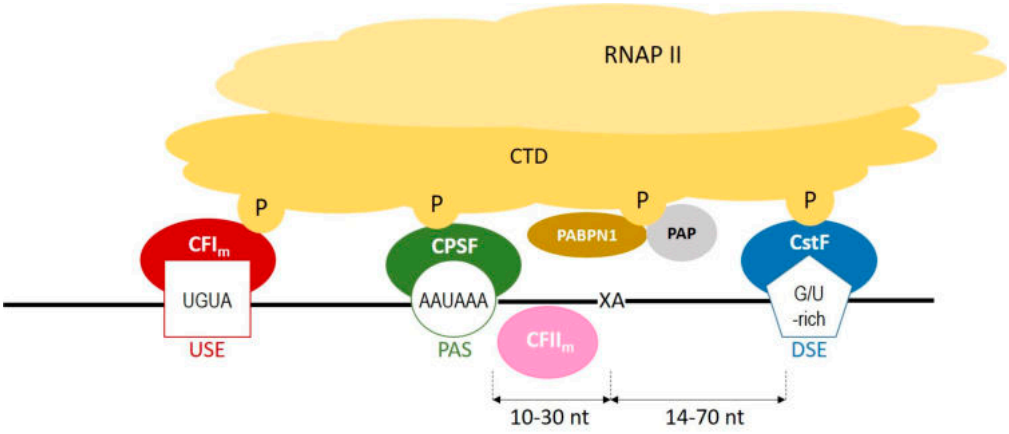


Figure 1. A scheme representing the cooperation between *cis*-elements and *trans*-factors, which are involved in a polyadenylation process. CFIm–cleavage factor Im; CFIIm–cleavage factor IIm; CPSF–cleavage and polyadenylation specificity factor; CstF–cleavage stimulation factor; CTD–carboxyterminal domain of RNA polymerase II; DSE–downstream element; P–phosphorylated serines of CTD; PABPN1–nuclear poly(A) binding protein; PAP–poly(A) polymerase; PAS–poly(A) signal; RNAPII–RNA polymerase II; USE–upstream element; XA–cleavage site.

2.2. Polyadenylation Roles

The poly(A) tail plays an important role in a transcript's life cycle [2]. For example, it is responsible for mediating the transport of mature mRNA into the cytoplasm through the NXF1-dependent pathway using nuclear-pore complexes embedded in the nuclear envelope [14]. Poly(A) tails are also involved in maintaining the stability of transcripts. They participate in the regulation of mRNA degradation through the process of deadenylation, which is commonly activated in post-transcriptional regulation via the sequence-specific binding of microRNA (miRNA), together with miRNA-induced silencing complex (miRISC), to the 3′-UTR [15,16]. Poly(A) tails are also important for translation regulation [17,18]. It is suggested that they interact functionally and physically with the 5′ cap of mRNA, creating a closed-loop structure that promotes the initiation of translation.

In metazoa, almost all mRNAs undergo polyadenylation. One of the exceptions is replication-dependent histone protein mRNAs, which form a highly conserved stem-loop structure [19,20]. Additionally, according to a recent comprehensive study [21], nearly 50% of long non-coding RNAs (lncRNAs) undergo polyadenylation, and the resulting poly(A) tails are important for their regulation at the cellular level [22].

2.3. Length and Composition of Tails

The lengths of poly(A) tails can vary between transcripts. For example, mRNAs of highly expressed genes, such as housekeeping genes, usually possess shorter poly(A) tails, whereas poorly translated transcripts and lncRNAs have longer tails [23,24]. Generally, the length range for specifying a short or long tail is dependent on the population of transcripts analyzed. For example, in a *Caenorhabditis elegans* study, half of the transcripts possessed poly(A) tails in the range of 70 to 94 A residues; therefore, short tails were defined as those with $\leq$70 A residues, and long ones, as those with >94 A residues [23]. The existence of shorter tails may be also explained by the fact that some genes (e.g., albumin and transferrin) contain poly(A)-limiting element (PLE), which tends to restrict the initial length of poly(A) tails from the pre-mRNA stage [25,26]. Additionally, a cytoplasmic polyadenylation element (CPE) can modulate the length of a poly(A) tail after the export of the transcript from the nucleus [27]. It is located in the 3′-UTR, near PAS, and its most common sequence is UUUUAU [28]. As investigated in neurons and during early development, CPE can influence the length of the poly(A) tail through the binding of proteins to this motif (e.g., CPEB) [29]. CPEB can promote either cytoplasmic polyadenylation or deadenylation. When it is deactivated (dephosphorylated), it recruits poly(A) ribonuclease deadenylase (PARN) to deadenylate and repress the expression of mRNA. On the other hand, when it is phosphorylated, it promotes the expulsion of PARN, thereby leading to polyadenylation by germ-line development factor 2 (GLD2) [26,30]. This cytoplasmic polyadenylation process occurs in mRNAs that already contain a short poly(A) tail and usually activates translation, leading to increased protein expression.

An interesting fact is that poly(A) tails are not limited to possessing A residues: reports indicate that cytosines, guanosines, and uridines can also be incorporated into the poly(A) tail. While the role of cytosines added to the poly(A) tail remains to be elucidated, the guanylation and uridylation of poly(A) tails are quite well understood. Guanylation occurs only for longer poly(A) tails and can slow down deadenylation, delaying transcript decay [31]. Guanylation frequently occurs for long-lived transcripts with a slow turnover, e.g., transcripts encoding secreted proteins [31]. On the other hand, uridylation is usually found in short tails and marks transcripts for decay [32,33]. This process tends to be a key factor in germline development, differentiation, and early embryogenesis, where short-lived transcripts, with relatively fast turnover, predominate [33]. The lengths of poly(A) tails might also depend on circadian rhythms and the cell cycle [34–36]. By studying multiple mouse liver mRNAs, researchers demonstrated that rhythmic changes in poly(A) tail lengths were under the control of the circadian clock. Even more importantly, they presented data indicating that rhythmic poly(A) tail lengths are correlated with rhythmic protein expression [34]. With regard to the cell cycle, TAIL-Seq analysis suggests that

global RNA decay takes place during the S phase through the accumulation of terminal uridylation. On the other hand, the accumulation of terminal guanylation occurs during the M phase of the cell cycle, leading to the assumption that the majority of the transcriptome is then protected from active deadenylation [35].

2.4. Alternative Polyadenylation (APA)

Apart from the canonical PAS-the AAUAAA hexamer-other weaker signals called alternative PASs, may be present in transcripts. Generally, the higher the sequence similarity between an alternative and canonical PAS, the stronger the recognition of the alternative PAS. When this alternative PAS is selected as a signal for the cleavage and polyadenylation event, the process is described as alternative polyadenylation (APA). APA is thought to occur for around 70% of human protein-coding genes and can also affect non-coding RNAs, such as lncRNAs [37–39]. The affected transcripts can exhibit various numbers of APA events in a few or multiple APA sites. APA can dramatically modulate the expression of a specific gene and affect the fate of its transcript, including its half-life and cellular localization [40,41]. Depending on the alternative PAS localization, APA can occur either in 3′-UTRs (UTR-APA) or upstream of the last exon: in introns or protein-coding exons (UR-APA). UR-APA can lead to the production of truncated proteins with different functions (protein diversification), or the production of dysfunctional proteins. On the other hand, when APA occurs in the 3′-UTR of a transcript, it leads to the creation of an mRNA of different lengths, which still codes for a full-length protein. In such cases, APA can affect the expression of a gene by, for example, changing the number of miRNA-binding sites in the transcript. As it was shown that the 3′-UTR can regulate protein localization independently from mRNA localization, it can act as a scaffold for various protein complexes which, when recruited to translation sites, can interact with specific domains of newly translated proteins [42]. This, in turn, leads to the translocation of such proteins. An example is a CD47 transcript, whose short 3′-UTR promotes the localization of the protein at the ER, while its longer isoform promotes its translocation to the plasma membrane [37]. The occurrence of APA can be regulated in many ways, one being the 'strength' of alternative PASs. The more similar the sequence of an alternative PAS is to that of the canonical PAS, the stronger the alternative PAS will be. Moreover, the localization of a specific PAS within a transcript sequence is also worth mentioning. Typically, PASs localized closer to the start codon (proximal) are considered to be weaker, while PASs localized further from the start codon (distal) are stronger [5]. Core polyadenylation factors, as well as other RNA-binding proteins (RBPs), can also regulate APA. For example, PABPN1 enhances the selection of distal PASs by competing with cleavage and polyadenylation complexes: it recognizes a weak PAS and binds to it, thereby blocking CPSF binding [43,44]. Regarding RBPs, HuR protein favors the selection of a distal PAS by binding to U-rich elements lying close to a proximal PAS [45]. Another crucial group of RBPs involved in APA is muscleblind-like (MBNL) proteins, whose binding sites are present in the close vicinity of many PASs. In myotonic dystrophy (DM), MBNL proteins were shown to either activate or suppress polyadenylation at specific sites [46]. PABPN1, MBNL, and HuR proteins are described in more detail in the following chapters.

3. Methods for Determining the Length of Poly(A) Tails and Analyzing APA Sites

Here, we briefly summarize some of the available methods for measuring poly(A) tail lengths using both transcript-specific and transcriptome-wide approaches, as well as methods designed for APA site identification and the currently available APA databases.

3.1. Transcript-Specific Analyses of Poly(A) Tails

For transcript-specific analyses, RNase H/Oligo(dT) northern blotting was once the gold-standard technique. Oligo(dT) were hybridized with transcripts, followed by RNase H treatment. Next, poly(A) tail length was assessed through the comparison of transcripts with, or without, hybridized oligo(dT) using northern-blot-based detection [47]. One

advantage of this technique was that it avoided introducing a PCR-based bias. However, it had many disadvantages, one being the requirement of large amounts of RNA. This technique is also laborious and works best on highly expressed and shorter mRNAs. These disadvantages encouraged scientists to develop more effective and more precise methods that allow for the transcript-specific measurement of poly(A) tails. Such techniques are usually grouped together and described as poly(A) test (PAT) assays, which include PCR-based rapid amplification of cDNA ends PAT (RACE-PAT), ligation-mediated PAT (LM-PAT) [48], and extension PAT (ePAT) [49]. These techniques, as well as other assays, were reviewed extensively in [26,50]. Currently, the most popular assay is probably ePAT. Here, a 5′-anchored oligo(dT) primer serves as a template for Klenow polymerase, which extends the transcript on the basis of the anchor sequence. This oligo(dT) then serves as a reverse primer for cDNA synthesis. When cDNA is generated, this technique requires the use of a target-specific forward primer for PCR. The PCR products can then be visualized using agarose gel electrophoresis or capillary electrophoresis. Another variation of the ePAT method is the poly(A) tail-length assay, currently available as a commercial kit [51]. This assay uses PAP, which adds guanosine and inosine residues at the 3′ end of every poly(A) tail that exists in the RNA sample. This added G/I tail then becomes a priming site for cDNA synthesis and the following PCR. These PAT methods are fast and easy to perform but require the optimization of the PCR. Otherwise, they may generate PCR-introduced biases and shorter products (i.e., those containing shorter poly(A) tails) may be amplified more efficiently.

3.2. Transcriptome-Wide Methods for Measurement of Poly(A) Tails

One method that can determine the lengths of poly(A) tails in both transcript-specific and transcriptome-wide manners is Nanopore technology sequencing [19,52]. Nanopore technology offers a set of products that allow for DNA and RNA sequencing, including for transcripts containing poly(A) tails. The principle of all Nanopore sequencing is that specific adaptors are attached to RNA or DNA, which then allow for such nucleic acids to be pulled through the pore at a constant rate. The length of a poly(A) tail is then estimated based on the correlation between length and the time spent being transferred through the pore. Thanks to numerous approaches, users can decide whether they want to use PCR amplification or bypass this step (to reduce PCR-introduced bias) and sequence the RNA or cDNA with only minimal library preparation, etc. This method is currently one of the best for measuring poly(A) tails and allows for the precise assignment of an analyzed poly(A) tail to a specific transcript. However, it requires large amounts of purified RNA and cannot identify non-A residues within the tail.

Other very popular methods for the transcriptome-wide measurement of poly(A) tail lengths include TAIL-Seq [53] and PAL-Seq [24]. Both methods are very similar in terms of their basic concepts: in the first step, RNAs lacking poly(A) tails (mainly rRNAs) are removed from a pool of sequenced transcripts. Moreover, biotinylated 3′ adaptors are used to allow for the pull-down of bound transcripts on streptavidin beads, which is followed by the addition of 5′ adaptors in both techniques. The main difference between TAIL-Seq and PAL-Seq is their sequencing methodology, which leads to slightly different results. TAIL-Seq allows for the determination of any non-A residues that might be present in the poly(A) tail (i.e., possible guanylation or uridylation). However, this technique is expensive and technically challenging. On the other hand, PAL-Seq permits the capture of tails with only adenosines present at their 3′ ends, but requires less starting RNA material and, due to the enrichment of polyadenylated RNAs at the start of the procedure, is more efficient than the TAIL-Seq approach [19]. PAL-Seq also requires a highly specialized experimental setup and can only be performed on the Illumina platform. In light of these disadvantages, another method that combines the enrichment of polyadenylated RNA from PAL-Seq with the sequencing methodology and algorithm of the TAIL-Seq protocol was developed. This method was named mTAIL-Seq [54] and requires even less starting RNA than the two previously discussed assays. This method also provides higher sensitivity and is less costly than the previous methods; however, it has a limited ability to capture tails shorter

than 8 nt and tails that end with non-A nucleotides. Due to its drawbacks, mTAIL-Seq is predominantly used for measuring poly(A) tail lengths.

Recently, a new technique has been developed: full-length poly(A) and mRNA sequencing (FLAM-Seq) [55]. FLAM-Seq relies on 3′ end RNA extension through G/I tailing, template switching, cDNA amplification, and PCR amplification. The transcript sequences are then determined through PacBio sequencing. It is a fast and simple method that allows for the generation of long reads and provides information on poly(A) tail length, sequences of multiple transcripts, and mRNA isoforms. An outstanding advantage of FLAM-Seq is that it allows for the sequencing of poly(A) tails with the detection of non-A residues inside the tail sequences. Moreover, it generates full-length sequences for thousands of mRNAs with a very low error rate [55].

3.3. APA-Identification Techniques and Databases

In discussing APA-identification techniques, we can distinguish between experimental methods and the various computational tools available (Table 1). An example of the former is a massive analysis of cDNA ends (MACE-Seq) [56]. This is an improved version of 3′ single-end mRNA-Seq, based on the sequencing of a single molecule of transcripts. Other experimental methods facilitating APA detection include TAIL-Seq, PAL-Seq, and FLAM-Seq, all of which were described in the previous sub-chapter. Regarding computational APA-identification methods, APAtrap allows for the identification and quantification of APA sites derived from various RNA-Seq data [57]. APAtrap also allows for the identification of 3′-UTRs and extended 3′-UTRs, as well as for distinguishing genes that use various APA sites under different conditions. Other computational tools for detecting APA, such as DaPars [58], scDAPA [59], etc., were previously reviewed in detail by others [11,60].

Table 1. Summary of experimental methods and computational tools allowing for APA sites detection together with APA databases (accessed on 12 February 2022).

Name	Year of Publication	Website	References
Computational tools for APA detection			
DaPars	2014	https://github.com/ZhengXia/dapars	[58]
Change point	2014	http://utr.sourceforge.net	[63]
Roar	2016	https://github.com/vodkatad/roar	[64]
APAtrap	2018	https://apatrap.sourceforge.io	[57]
QAPA	2018	https://www.github.com/morrislab/qapa	[65]
TAPAS	2018	https://github.com/arefeen/TAPAS	[66]
KAPAC	2018	https://github.com/zavolanlab/PAQR_KAPAC.git	[67]
scDAPA	2019	https://scdapa.sourceforge.io	[59]
APA databases			
PolyA-Seq Atlas	2012	http://genome.ucsc.edu/	[10]
APADB	2014	http://tools.genxpro.net:9000/apadb/	[62]
APASdb	2015	http://genome.bucm.edu.cn/utr/	[68]
PolyA_DB3	2018	https://exon.apps.wistar.org/PolyA_DB/v3/	[69]
APAatlas	2020	https://hanlab.uth.edu/apa/	[70]
PolyAsite 2.0	2020	https://polyasite.unibas.ch	[61]
Animal-APAdb	2021	http://gong_lab.hzau.edu.cn/Animal-APAdb/	[71]

Over recent years, several databases that gather information about possible APA sites in various transcripts have been created. The data collected there were obtained using the previously mentioned RNA-Seq methods and computational tools. One such database is PolyASite 2.0 [61]. It contains information about PASs based on data generated using numerous sequencing methods for human, mouse, and *C. elegans* transcriptomes [61]. Moreover, PolyASite 2.0 provides information about PASs located in 3′-UTRs, introns, and CDSs. Another database is APADB [62], which contains information about human, mouse, and chicken transcripts of both coding and non-coding genes. The data collected there were generated through the NGS-coupled 3′ end sequencing of thousands of samples (e.g., MACE-Seq). APADB also provides information regarding the loss of bioinformatically predicted miRNA-binding sites, which may occur through APA events themselves. Other databases were well-reviewed recently [11,60].

4. Repeat Expansion Diseases

Repeat expansion diseases are a group of genetic disorders caused by expansions of polymorphic tracts of nucleotide sequences in specific human genes [72] (Table S1). There are more than 40 diseases that are classified as repeat expansion disorders, and this number continues to grow [72,73]. Most of them mainly affect the nervous system, but there are also some diseases that impact the muscular system. The most common expanded sequence motifs are trinucleotides, but longer ones, up to dodecanucleotide units, exist [74]. Repeat expansion diseases display several specific features: they are mostly autosomal dominant diseases, but recent reports indicate that mutations of expanded repeats also cause autosomal recessive diseases, e.g., cerebellar ataxia, neuropathy, and vestibular areflexia syndrome (CANVAS) [73]. Typically, patients are heterozygotes possessing wild-type and mutant alleles, with the latter harboring the expanded repeat tract. With regard to the mutant allele, usually the longer the repeat tract, the earlier the onset of symptoms and the more severe the disease course. However, the number of nucleotide repeats needed for disease onset differs among diseases. Expanded repeat tracts can occur in protein-coding sequences of particular genes, as well as in promoter sequences, introns, or in 5′- and 3′-UTRs (Table S1). Interestingly, many expanded repeat tracts in RNAs are prone to the formation of stable structures like hairpins or G-quadruplexes [75,76]. These structures affect transcript functioning, e.g., through sequestration of proteins [72], whereas many processes remain to be investigated in this context, including APA.

In this chapter, we shortly summarize the tissue-specific processing of poly(A) tails, focusing on the cell types most affected in repeat expansion diseases. Moreover, we discuss selected diseases in the context of APA and poly(A) tail lengths in the implicated mutant transcripts.

4.1. APA and Poly(A) Tail Lengths in Neurons and Muscles

An interesting aspect concerning polyadenylation in neurons is the occurrence of neuron-specific proteins that demonstrate a specific impact on polyadenylation. These proteins include the previously mentioned GLD2 [77], mammalian suppressor of tauopathy 2 (MSUT2) [78], zinc finger CCCH domain-containing protein 14 (ZC3H14) [79], and ataxin-2 (ATXN2) [80]. In terms of the present review, the most interesting protein is ATXN2, as the expanded repeat tract in *ATXN2* is responsible for spinocerebellar ataxia type 2 (SCA2) [81], as well as normal ATXN2 is implicated in amyotrophic lateral sclerosis (ALS) [82]. Regarding polyadenylation, ATXN2 inhibits poly(A) nuclease (PAN) activity and causes the occurrence of longer poly(A) tails [80]. It was also suggested that ATXN2 might stabilize its associated mRNAs by suppressing their decay. Additionally, it may contribute to the APA of its associating mRNAs, thereby supporting activity-dependent translation and neural plasticity [80].

Over recent years, it has been shown that, for many genes, the selection of a particular polyadenylation site in a transcript is tissue-specific [83–85]. Regarding neurons, studies performed in *Drosophila*, mouse, and human cells demonstrated a neuron-specific enrichment of long 3′-UTRs in various transcripts [41,86–88]. An example of such a phenomenon is *Drosophila's* mei-P26 transcript, whose 3′-UTR is very short in the testis and 18.5 kb long in the central nervous system (CNS) [86]. A tendency for longer 3′-UTR expression in the CNS was suggested to be limited to neurons and is not the case in astrocytes, microglia, or oligodendrocytes [58,63]. An example of a factor that favors the choice of distal PASs is the neuron-specific RBP Elav (embryonic lethal abnormal visual protein), first identified in *Drosophila* [88,89]. Elav binds to its target mRNA at proximal PASs, blocking cleavage and polyadenylation machinery, thereby promoting the use of distal PASs. In *Drosophila* embryos, in the absence of the Elav protein, several genes lack long 3′-UTRs [90]. Human homologs of the Drosophila Elav protein are the Hu-family proteins: HuR, HuB, HuC, and HuD [91].

Due to neuron-specific features such as their structure (a single axon can extend hundreds of centimeters in length) and function (the synaptic plasticity underlying learning and memory processes), a common feature of these cells is local translation [17,30,92]. This phenomenon is required at synapses to, for example, convert short-term memory

into long-term memory. Therefore, neurons have had to develop specific mechanisms for the transport and repression of the translation of transcripts (during their transport) throughout the cell, to its distal compartments (e.g., dendrites) [17,92]. Here, the α-subunit of calcium/calmodulin-dependent protein kinase II (αCaMKII) transcript serves as an example (reviewed in [17]). When needed in dendrites, this transcript is polyadenylated in the nucleus and exported to the soma, where it is deadenylated and repressed. It is transported in this state to dendrites and, upon neuronal stimulation, is polyadenylated again and locally translated. This is also an example of the regulation of poly(A) tail length in neurons, which depends on the cell's needs.

Regarding the muscular system, RNA-sequencing studies performed in *C. elegans* demonstrated the general use of APA sites in muscles. This research revealed that muscle tissue demonstrates a specific polyadenylation pattern [93]. An example of such a tissue-specific form of post-transcriptional regulation led by APA is the *PAX3* (Paired box gene-3) gene, which is a major regulator of muscle stem-cell development and is required for the formation of limb muscles [94] and in response to stress [95,96]. During myogenic differentiation, *PAX3* expression is silenced by miR-206 [97]. The APA-based use of proximal PAS in myogenic progenitors leads to 3′-UTR shortening of the *PAX3* transcript, which results in such mRNA escaping miRNA-induced gene silencing [97–99]. In muscles, RBPs also play an important role in the control of PAS selection and have an impact on the general polyadenylation profile. Among such RBPs we specify MBNL proteins, which are discussed in detail in the DM 1 subsection below.

4.2. Characteristics and Perturbations in Poly(A) Tail Processing in Repeat Expansion Diseases

4.2.1. Oculopharyngeal Muscular Dystrophy

A repeat expansion disease with crucial APA implications is dominant oculopharyngeal muscular dystrophy (OPMD), which is characterized by progressive eyelid drooping (ptosis), difficulty in swallowing, filamentous intranuclear inclusions in muscle fibers, and proximal limb weakness [100,101]. The gene encoding PABPN1, mentioned in the first chapter, is directly involved in this disease. The relevant mutation involves the expansion of a GCG repeat tract encoding the polyalanine tract, located at the *N*-terminus of the PABPN1. This protein plays an important role in RNA processing, e.g., it is required for efficient mRNA export from the nucleus [102] and controls the length of poly(A) tails [13]. The mutated version of PABPN1 was shown to sequester the WT protein in nuclear inclusions in the muscle fibers of OPMD patients, altering the function of the WT protein [103]. Moreover, it was recently suggested that the PABPN1 may act as an APA suppressor. This hypothesis was supported by the finding that this protein binds near proximal PASs and therefore suppresses the use of these PASs for cleavage and polyadenylation. Additionally, it was demonstrated that either the downregulation or mutation of PABPN1 resulted in the genome-wide use of proximal PASs, which finally cause 3′-UTR shortening to occur in both mouse models of OPMD and in human cells [43,44].

4.2.2. Fragile X-Associated Tremor/Ataxia Syndrome

Fragile X syndrome (FXS), and its related diseases-fragile X-associated immature ovarian insufficiency (FXPOI) and fragile X-associated tremor/ataxia syndrome (FXTAS)-are caused by the expansion of a CGG repeat tract located in the 5′-UTR of *fragile X mental retardation 1* gene (*FMR1*) [104]. In FXPOI and FXTAS, premutation alleles contain 55–200 CGG repeats; however, this mutation does not inactivate the gene [104]. On the other hand, in FXS, the full mutation in *FMR1* of more than 200 CGG repeats leads to transcriptional silencing of the gene [104]. Normally, *FMR1* encodes a protein (FMRP) which is important in brain development as it plays a role in the functioning of neuronal networks and neuronal plasticity [105,106]. In the 3′-UTR of an *FMR1* transcript, three PASs were identified: one canonical and two alternatives [107]. In premutation alleles, the level of transcript isoforms emerging from alternative PASs is decreased, indicating that *FMR1* premutation affects polyadenylation through APA phenomena and thus might

impact pathology. Additionally, it was shown that, in this case, APA produces mRNAs with different poly(A) tail lengths: a transcript with expanded CGG tract had a shorter poly(A) tail than its WT counterpart [107].

4.2.3. Amyotrophic Lateral Sclerosis

ALS is a neurodegenerative disease that causes the death of motor neurons that control skeletal muscle contraction. ALS progresses over time, leading to muscle stiffness, weakness, and finally, paralysis and death. This process takes 2–5 years from the initial appearance of symptoms. ALS can have a sporadic or familial character. To date, the pathogenesis of sporadic cases remains unclear, while familial ones are mainly caused by mutations in *C9orf72*, *SOD1*, *FUS*, or *TARDBP* [108,109].

The most common familial form of ALS is caused by the expansion of GGGGCC repeats in the first intron of the *C9orf72* gene. This intron should be degraded after splicing, but because of the mutation, it accumulates in the nucleus and forms RNA foci [110]. Alternatively, it is transported to the cytoplasm where it undergoes RAN translation which, in turn, creates potentially toxic polypeptides [108,111]. RNA-seq was used to determine the APA status of samples from ALS patients, and the results indicated that many defects exist in the overall polyadenylation of transcripts in ALS. Depending on the brain area and type of ALS, distal or proximal PASs are used: (I) in the case of the cerebellum in familial ALS, proximal, rather than distal, PASs are used, while in sporadic ALS, the use of proximal and distal PASs is almost equivalent; (II) in the frontal cortex in familial ALS, proximal and distal PASs have similar rates of use, while in sporadic ALS, distal PASs are used more than in proximal ones [112]. Similar studies demonstrated that 2.7% of genes have statistically significant deregulation of APA in ALS, with a tendency toward the extension of 3′-UTRs. Gene ontology analysis revealed that most of these genes are responsible for neuron-projection development and cytoskeletal intracellular transport [109]. In ALS, the deregulation of polyadenylation affects many genes and potentially seems to be part of the pathogenesis mechanism.

Other genes whose mutations cause ALS, *FUS*, and *TARDBP*, encode two RBPs: FUS and TDP-43 respectively. In *TARDBP*, the mutations can be localized in different regions, but the majority are in disordered low-complexity regions [113,114]. Moreover, in *FUS*, other mutations that cause ALS have been identified, with the most common being a substitution of R521 [114]. The localization of FUS and TDP-43 mutations in their nuclear localization signals (NLS) causes the depletion of these proteins from the nucleus and leads to their accumulation in inclusions in the cytoplasm. Both proteins control the alternative splicing of genes participating in neuronal development which, in the case of mutation, leads to neurodegeneration [115]. Moreover, FUS and TDP-43 are involved in the polyadenylation process. FUS can control the selection of the polyadenylation site depending on the distance from the FUS-binding site [115]. TDP-43′s role is to repress cryptic exons, thereby stabilizing transcripts. TDP-43 suppresses a cryptic PAS of the *stathmin-2* mRNA, which encodes proteins crucial for the regeneration and growth of axons. The re-localization or depletion of TDP-43 causes the recognition of a premature PAS in the *stathmin-2* mRNA and leads to the production of non-functional transcripts [114]. TDP-43 also controls the PAS selection of its own transcript and, in this way, self-regulates its expression. In the case of an excess of TDP-43 in the nucleus, it inhibits the selection of the most proximal PAS (there are three alternative PASs within the TARDBP sequence), leading to the elongation of a transcript. In such cases, the splicing of additional introns occurs, which is followed by the export of the transcript to the cytoplasm and its degradation by nonsense-mediated mRNA decay. As a result, a decrease in the cytoplasmic level of *TARDBP* mRNA is observed. On the other hand, when there is less nuclear TDP-43, the proximal PAS is selected. In this case, the transcript is transported to the cytoplasm and translated. In ALS patients, mutant TDP-43 is depleted from the nucleus, which leads to an increase in the level of TDP-43 in motor neurons by the mechanism described above. This loop may accelerate the progress of the disease through the formation of cytoplasmic inclusions and disturbances in mRNA metabolism [116].

4.2.4. Myotonic Dystrophy Type 1

DM 1 is primarily characterized by muscle dystrophy or myotonia. A mutation responsible for DM1 is the expansion of CUG repeats present in the 3′-UTR of the *DMPK* gene. RNA containing the expanded CUG tract acts through gain of function mechanisms. Mutated transcripts are mainly localized in the nucleus, where they form RNA foci and sequester RBPs such as MBNL1, -2, and -3; HNRNPH1; CUGBP1; and STAU1 [117,118]. This leads to the trans-deregulation of RNA metabolism-especially APA, miRNA processing, and alternative splicing processes [119,120]. MBNL's implications in alternative splicing and APA are well-documented [121–124]. The RNA-binding motif of this protein is composed of four zinc-finger domains [125]. Using this motif, MBNL binds to transcripts and act as an activator or repressor of splicing [126]. It was shown that MBNL1 can regulate alternative splicing in the brain, heart, and muscle in post-natal development [46,127]. In DM1, the sequestration of MBNL1 results in the functional inactivation of the protein, leading to disturbances in the alternative splicing of genes controlled by this protein [128,129]. Concerning APA, in MEF cells with *Mbnl1*, *Mbnl2*, and *Mbnl3* depletion, thousands of polyadenylation sites shifts were detected, which suggests that MBNL proteins act globally on alternative translation events [46]. The importance of MBNL1 sequestration in DM1 pathogenesis was demonstrated in skeletal muscle satellite cells [130]. These cells are responsible for muscle regeneration, and, in cases of DM1, this regeneration is impaired. Moreover, autophagy is increased, lowering the proliferation ability of skeletal satellite cells-a process that plays a key role in the initial steps of cell regeneration in damaged muscle. However, the genome modification of *DMPK*, or overexpression of the MBNL1 protein, can ameliorate proliferation defects in skeletal-muscle satellite cells. Indeed, it was revealed that increases in MBNL1 levels in the cytoplasm partially restored appropriate RNA processing [130].

4.2.5. Huntington's Disease

Huntington's disease (HD) is a neurodegenerative disorder caused by the expansion of the CAG repeat tract, encoding the polyglutamine (polyQ) tract, in the first exon of the huntingtin gene (*HTT*) [131,132]. The expansion is dominantly inherited, and patients usually develop symptoms in their mid-thirties. Patients harbor two alleles: the wild-type, containing 10 to 30 CAG repeats, and the mutant allele, with over 40 CAG repeats. The mutation of the CAG tract in *HTT* leads to the degeneration of neurons, which emerges in the motor cortex and striatum and is also substantial in other regions of the brain [132,133]. Huntingtin transcripts were reported to have several isoforms produced via APA. Three of these APA events occur in the 3′-UTR of the *HTT* gene, leading to three isoforms that differ only in their 3′-UTR length, with no impact on the coding sequence [134] (Figure 2a, three arrows marked in 3′-UTR). These 3′-UTR isoforms of the *HTT* transcript include a short isoform (PAS: AGUAAA), which is 10.3 kb long, and a long isoform (PAS: AUUAAA), which is 13.7 kb long [134,135]. The third one, 12.5 kb long, is described as an intermediate 3′-UTR isoform (PAS: AAUGAA) and, as well as the previous isoforms, is conserved between mice and humans [134]. This research also suggested that all of these *HTT* isoforms may have different half-lives and localization, as well as RBP- and miRNA-binding sites. Short and long 3′-UTR isoforms are considered to be the most common *HTT* isoforms: the short one is more abundant in actively dividing cells (e.g., B cells and the HEK 293 cell line) and peripheral tissues (muscles), while the long one predominates in non-dividing cells (i.e., terminally differentiated cells), such as in brain tissue, but also prevails in breast and ovary tissues [134,135]. Research using SH-SY5Y cells and focusing on poly(A) tail length in *HTT* transcripts revealed that each of these three *HTT* mRNA isoforms had different lengths of poly(A) tails. The short 3′-UTR isoform possesses around 60 adenosine residues in its poly(A) tail, while the intermediate and long 3′-UTR isoforms have around 5 and 10 A residues, respectively [134]. These results suggested that *HTT* transcripts have rather short poly(A) tail lengths, at least in this specific cell line. It was also shown that these isoforms arise from both alleles (normal and mutant) in HD, meaning that CAG expansion

probably does not have an impact on PAS selection in the 3′-UTR of *HTT* [134]. However, another study concluded that the short 3′-UTR isoform might be translated more efficiently than the longer one [136].

In HD, an additional isoform of the transcript (*HTTexon1)* is produced predominantly from mutant *HTT* due to APA site selection in intron 1 [137] (Figure 2a, red arrow in the first intron of *HTT*). This APA event results in an isoform of 7.9 kb, which is translated into the highly truncated huntingtin protein, with a high tendency to form aggregates. This mRNA isoform has been detected in HD patients but not in healthy individuals [137]. Another study revealed higher levels of this isoform in the HD patient hippocampus and motor cortex compared with controls [138]. Those results strongly suggest that the production of *HTT* mRNA containing exon 1 and part of intron 1 is associated with HD. One hypothesis that might explain this generation of such an mRNA isoform is related to the serine/arginine-rich splicing factor 6 (SRSF6). Bioinformatic analysis predicted that SRSF6 binds to CAG and CAGCAAA repeats, resulting in the incomplete splicing of exon 1 to exon 2 (intron 1 retention) and simultaneously leading to the exposure of cryptic poly(A) sites within intron 1 [139]. However, a subsequent in vivo study performed by the same group demonstrated that SRSF6 is not needed for the aberrant splicing of *HTT* [140], meaning that this protein and its sequestration by CAG repeats is not responsible for a mechanism leading to the formation of *HTTexon1* mRNA, which warrants further investigation.

4.3. Potential Perturbations in Poly(A) Tail Processing in Other PolyQ Diseases

HD is one of nine polyQ diseases that are all caused by CAG repeat expansion located in the ORF regions of specific genes [141]. Due to already-known perturbations in poly(A) tail processing in *HTT*, as described in the previous subchapter, we more closely examined other transcripts directly involved in polyQ diseases. Apart from *HTT*, we searched the PolyASite 2.0 database [61] for information concerning alternative polyadenylation sites in eight other mRNAs: *AR*, *ATN1*, *ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*, and *TBP*, whose genes contain mutations responsible for spinal and bulbar muscular atrophy (SBMA), dentatorubral-pallidoluysian atrophy (DRPLA), and spinocerebellar ataxia types 1, 2, 3, 6, 7, and 17 (SCA1, SCA2, SCA3, SCA6, SCA7, and SCA17, respectively). For all the mRNAs, 3′ end sequencing data were available and included a set of sites for alternative polyadenylation (Figure 2b–i). Apart from 3′-UTRs, identified PASs were located in protein-coding exons and introns, but the use of these sites was often quantified as very low. Nevertheless, the identification of these sites suggests the possibility of premature transcription termination (PTT) [142] for these transcripts which may specifically occur under certain conditions. Selected sites are presented in Figure 2, where we include a presentation of PASs with relatively high TPM values ($\geq$0.1) within all the gene regions, as well as all the PASs that fulfilled at least one of the following criteria: (I) presence in the same exon as a CAG repeat tract, (II) presence in an exonic sequence at a distance of less than 500 nt from a CAG repeat tract, or (III) presence in introns directly adjacent to an exon containing CAG repeat tract. The purpose of these criteria was to verify the presence of PASs that could potentially be of importance in cases of repeat-tract mutations. Indeed, numerous PASs were identified according to the listed criteria for most analyzed mRNAs. Abnormal interactions of proteins with expanded CAG repeats in RNA, or the abnormal secondary structure of this mutated region [143], can potentially lead to disruptions in splicing and/or PAS recognition. In several of the investigated transcripts, PASs were present in neighboring introns, which could lead to the production of a shorter transcript and smaller protein if a splicing aberration—similar to that reported for *HTT*—occurs. These effects may be crucial for pathogenesis, especially for mRNAs where the CAG repeat tract is located closer to the 5′ end of a transcript, such as in *HTT*. For mRNAs including *AR*, *ATXN2*, and *TBP*, if the resulting transcript would still contain the mutant CAG tract, it would be translated into a very short aberrant protein with expanded polyQ tracts that could be very prone to aggregation. These events are speculative at present, but, as was shown in HD, may be important contributors in cascades of pathogenic events.

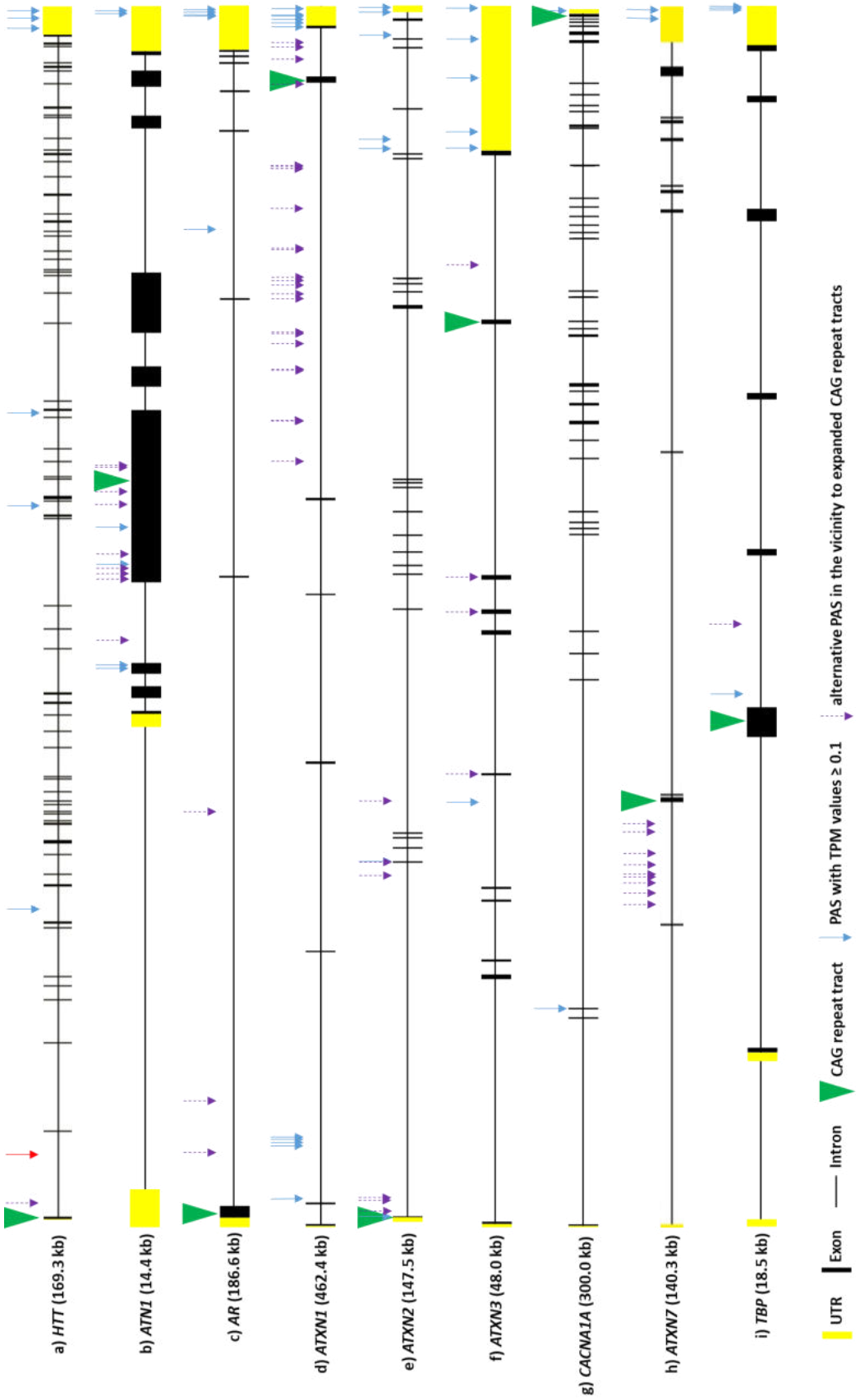


Figure 2. A scheme representing genes associated with polyQ diseases, based on Ensembl data [144], with marked alternative polyadenylation sites derived from PolyASite 2.0 database. The red arrow represents PAS identified in intron 1 of mutant *HTT* gene. See text for more details.

In a set of analyzed mRNAs, there exists variation in the number and distribution of PASs in the 3′-UTRs (Figure 2). Most of the transcripts have more than one PAS identified in the 3′-UTR. As previously mentioned, one of the consequences of the selection of proximal or distal PASs, especially in the 3′-UTR, is the presence or absence of miRNA-binding sites. We included additional analysis of the data gathered in APAdb [62] to list miRNAs that lose their binding sites in the case of proximal or intermediate PAS selection in *HTT* (Figure 3a) and *ATXN3* (Figure 3b) mRNAs. Numerous miRNAs are identified as potentially regulating these transcripts and potential proximal PAS selection will lead to a lack of these expression-regulating sites.

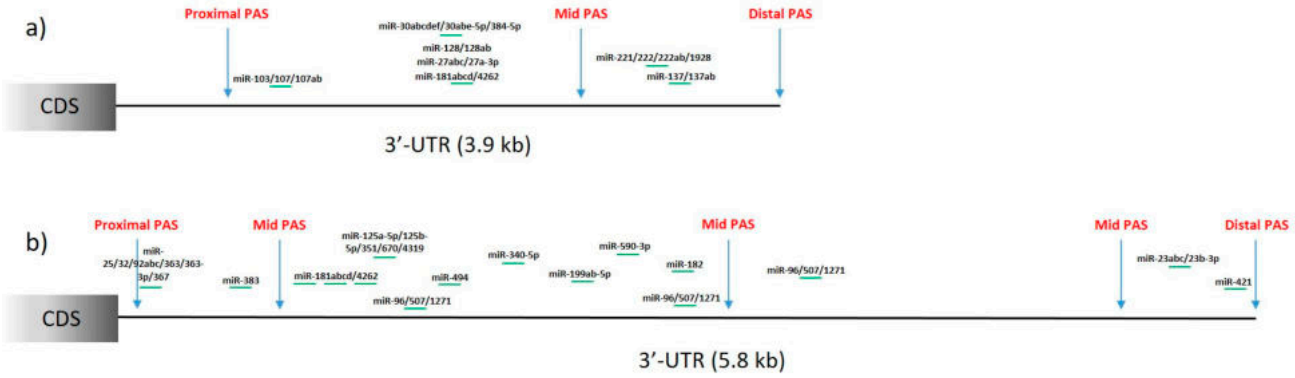


Figure 3. A schematic representation of miRNA-binding sites that are present in 3′-UTRs of *HTT* (**a**) and *ATXN3* (**b**) mRNAs with available PASs marked. Selection of PAS other than distal will lead to a lack of specific miRNA-binding sites.

5. Conclusions and Perspectives

3′-UTRs, where polyadenylation events usually occur, had their lengths generally increased during evolution and therefore are considered as one of the elements of increased molecular, and as a result also morphological, complexity [145]. APA adds another level for the regulatory role of 3′-UTRs, as transcripts that substantially differ in 3′-UTR length can be generated. This results in the presence, or absence, of specific structural elements, RNA- and protein-binding regions, etc., and thus it affects the metabolism of a given transcript [146].

APA has various biological functions, both physiological and pathological. The most recent data suggest that APA can be a molecular tool for regulating mRNA and protein levels during cellular responses, e.g., differentiation [38]. On the other hand, as APA can affect the expression of various genes and can produce incorrect proteins, it is also associated with many diseases, whether oncological, immunological, endocrine, or neurological.

5.1. Methodological Aspects

As poly(A) tail processing is involved in many physiological and pathological mechanisms, there is still a great need for the development of various tools and methods that will allow for more insightful studies of APA sites and tail lengths. It has now become possible to precisely analyze polyadenylation at the single-cell level [41,147,148], but data gathered from multiple cell types and at different developmental stages is needed to advance our understanding of events connected to polyadenylation. Moreover, this process is clearly associated with other molecular mechanisms that occur for specific transcripts, and a further consolidation of the available data is required to provide a comprehensive view of transcript functioning.

5.2. Implications for Pathogenesis of Repeat Expansion Diseases

As the majority of transcripts undergo the selection of APA sites, it remains to be elucidated how eventual mis-selection events are connected with diseases. In mutant

transcripts, repeat tract expansion can affect PAS selection, as was described in HD. Repeat expansion diseases mainly affect specific cell types, e.g., a specific vulnerability of neuronal cells to degeneration is observed [149]. Gathering more and more data from disease tissues and cells will enable the identification of potential disruptions which could occur in specific cell types in cases of repeat tract mutation. Additionally, as widespread APA deregulation has been already observed in a range of neurological disorders [109], it would be of advantage to establish what are the specifically deregulated APA events also in repeat expansion diseases. This knowledge could be used for molecular diagnostics, as it has been recently suggested that APA-signatures in cancer appear to outperform existing biomarkers [58,150,151].

5.3. Implications into Therapy of Repeat Expansion Diseases

All repeat expansion diseases are as yet incurable, and often, especially for diseases where the expanded repeat tract is located in a coding sequence, the most straightforward therapeutic approach would be to eliminate the expression of the mutant gene. Many challenges remain to be faced in the development of these strategies. A crucial aim is to advance our knowledge regarding the therapeutic target features that could allow for the design of the most efficient and safest molecules for therapy. Mutant transcripts are currently the most promising targets, as various tools can be designed to bind to them in order to inhibit mutant gene expression. In regards to the therapeutic aspects, it remains to be determined whether specific 3′ end transcript variants are present in cases of repeat tract mutation, and, if so, how they could be specifically targeted. Specific therapeutic interventions based on modulation of biogenesis or direct elimination of faulty APA isoforms are being currently extensively developed for other diseases such as cancer (as reviewed in [3]). Currently, the most desirable therapeutic tools for the majority of repeat expansion diseases are allele-selective ones, which preferentially target the expression of the mutant allele. Various types of oligonucleotides-including short interfering RNAs (siRNAs) and antisense oligonucleotides (ASOs)-have been tested in animal models, and clinical testing is ongoing [152]. Interestingly, some of the mutant mRNA-targeting oligonucleotides may cause deadenylation. Our recent study, focusing on the mechanisms leading to the allele-selective silencing of *HTT*, demonstrated that a specific self-duplexing siRNA (sd-siRNA) [153], which is described as miRNA-like siRNA, caused the rapid deadenylation of the mutant transcript and translational inhibition [154]. It remains to be investigated how deadenylation is activated in this therapeutic strategy using an endogene of *HTT*.

It seems we are greatly advancing our understanding of how important the poly(A) tail-connected events are in the molecular biology puzzles. In repeat expansion diseases, we already know some relationships of APA to selected diseases, but it looks like a lot is ahead of us to unravel.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cells11040677/s1, Table S1: Repeat expansion diseases.

Author Contributions: General conceptualization by A.F. and A.C.; P.J. wrote the majority of manuscript draft and prepared figures and tables. G.F., A.F. and A.C. wrote some fragments and provided comments for manuscript draft. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Centre (2015/19/B/NZ2/02453).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

ALS	amyotrophic lateral sclerosis
APA	alternative polyadenylation
DM	myotonic dystrophy
HD	Huntington's disease
OPMD	oculopharyngeal muscular dystrophy
PABPN1	poly(A) binding protein nuclear 1
PAS	polyadenylation signal
PAT	poly(A) test
polyQ diseases	polyglutamine diseases

References

1. Casañal, A.; Kumar, A.; Hill, C.H.; Easter, A.D.; Emsley, P.; Degliesposti, G.; Gordiyenko, Y.; Santhanam, B.; Wolf, J.; Wiederhold, K.; et al. Architecture of eukaryotic mRNA 3′-end processing machinery. *Science* **2017**, *358*, 1056–1059. [CrossRef] [PubMed]
2. Passmore, L.A.; Coller, J. Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. *Nat. Rev. Mol. Cell Biol.* **2021**. [CrossRef] [PubMed]
3. Nourse, J.; Spada, S.; Danckwardt, S. Emerging roles of RNA 3′-end cleavage and polyadenylation in pathogenesis, diagnosis and therapy of human disorders. *Biomolecules* **2020**, *10*, 915. [CrossRef] [PubMed]
4. Stewart, M. Polyadenylation and nuclear export of mRNAs. *J. Biol. Chem.* **2019**, *294*, 2977–2987. [CrossRef] [PubMed]
5. Di Giammartino, D.C.; Nishida, K.; Manley, J.L. Mechanisms and Consequences of Alternative Polyadenylation. *Mol. Cell* **2011**, *43*, 853–866. [CrossRef]
6. Sadek, J.; Omer, A.; Hall, D.; Ashour, K.; Gallouzi, I.E. Alternative polyadenylation and the stress response. *Wiley Interdiscip. Rev. RNA* **2019**, *10*, 1–16. [CrossRef]
7. Ren, F.; Zhang, N.; Zhang, L.; Miller, E.; Pu, J.J. Alternative Polyadenylation: A new frontier in post transcriptional regulation. *Biomark. Res.* **2020**, *8*, 67. [CrossRef]
8. Danckwardt, S.; Hentze, M.W.; Kulozik, A.E. 3′ end mRNA processing: Molecular mechanisms and implications for health and disease. *EMBO J.* **2008**, *27*, 482–498. [CrossRef]
9. Proudfoot, N.J. Ending the message: Poly(A) signals then and now. *Genes Dev.* **2011**, *25*, 1770–1782. [CrossRef]
10. Derti, A.; Garrett-Engele, P.; MacIsaac, K.D.; Stevens, R.C.; Sriram, S.; Chen, R.; Rohl, C.A.; Johnson, J.M.; Babak, T. A quantitative atlas of polyadenylation in five mammals. *Genome Res.* **2012**, *22*, 1173–1183. [CrossRef]
11. Gruber, A.J.; Zavolan, M. Alternative cleavage and polyadenylation in health and disease. *Nat. Rev. Genet.* **2019**, *20*, 599–614. [CrossRef] [PubMed]
12. Zhao, J.; Hyman, L.; Moore, C. Formation of mRNA 3′ Ends in Eukaryotes: Mechanism, Regulation, and Interrelationships with Other Steps in mRNA Synthesis. *Microbiol. Mol. Biol. Rev.* **1999**, *63*, 405–445. [CrossRef]
13. Eckmann, C.R.; Rammelt, C.; Wahle, E. Control of poly(A) tail length. *Wiley Interdiscip. Rev. RNA* **2011**, *2*, 348–361. [CrossRef] [PubMed]
14. Natalizio, B.J.; Wente, S.R. Postage for the messenger: Designating routes for nuclear mRNA export. *Trends Cell Biol.* **2013**, *23*, 365–373. [CrossRef] [PubMed]
15. Bresson, S.M.; Conrad, N.K. The Human Nuclear Poly(A)-Binding Protein Promotes RNA Hyperadenylation and Decay. *PLoS Genet.* **2013**, *9*, 1003893. [CrossRef]
16. Eisen, T.J.; Eichhorn, S.W.; Subtelny, A.O.; Lin, K.S.; McGeary, S.E.; Gupta, S.; Bartel, D.P. The Dynamics of Cytoplasmic mRNA Metabolism. *Mol. Cell* **2020**, *77*, 786–799. [CrossRef]
17. Weill, L.; Belloc, E.; Bava, F.-A.; Méndez, R. Translational control by changes in poly(A) tail length: Recycling mRNAs. *Nat. Struct. Mol. Biol.* **2012**, *19*, 577–585. [CrossRef]
18. Xiang, K.; Bartel, D.P. The molecular basis of coupling between poly(A)-tail length and translational efficiency. *eLife* **2021**, *10*, 66493. [CrossRef]
19. Nicholson, A.L.; Pasquinelli, A.E. Tales of Detailed Poly(A) Tails. *Trends Cell Biol.* **2019**, *29*, 191–200. [CrossRef]
20. Dominski, Z.; Yang, X.-C.; Kaygun, H.; Dadlez, M.; Marzluff, W.F. A 3′ Exonuclease that Specifically Interacts with the 3′ End of Histone mRNA. *Mol. Cell* **2003**, *12*, 295–305. [CrossRef]
21. Lorenzi, L.; Chiu, H.S.; Avila Cobos, F.; Gross, S.; Volders, P.J.; Cannoodt, R.; Nuytens, J.; Vanderheyden, K.; Anckaert, J.; Lefever, S.; et al. The RNA Atlas expands the catalog of human non-coding RNAs. *Nat. Biotechnol.* **2021**, *39*, 1453–1465. [CrossRef]
22. Chen, L.L. Linking Long Noncoding RNA Localization and Function. *Trends Biochem. Sci.* **2016**, *41*, 761–772. [CrossRef]
23. Lima, S.A.; Chipman, L.B.; Nicholson, A.L.; Chen, Y.H.; Yee, B.A.; Yeo, G.W.; Coller, J.; Pasquinelli, A.E. Short poly(A) tails are a conserved feature of highly expressed genes. *Nat. Struct. Mol. Biol.* **2017**, *24*, 1057–1063. [CrossRef]
24. Subtelny, A.O.; Eichhorn, S.W.; Chen, G.R.; Sive, H.; Bartel, D.P. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **2014**, *508*, 66–71. [CrossRef]
25. Gu, H.; Gupta, J.D.; Schoenberg, D.R. The poly(A)-limiting element is a conserved cis-acting sequence that regulates poly(A) tail length on nuclear pre-mRNAs. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 8943–8948. [CrossRef] [PubMed]

26. Jalkanen, A.L.; Coleman, S.J.; Wilusz, J. Determinants and implications of mRNA poly(A) tail size—Does this protein make my tail look big? *Semin. Cell Dev. Biol.* **2014**, *34*, 24–32. [CrossRef] [PubMed]
27. Charlesworth, A.; Meijer, H.A.; de Moor, C.H. Specificity factors in cytoplasmic polyadenylation. *Wiley Interdiscip. Rev. RNA* **2013**, *4*, 437. [CrossRef] [PubMed]
28. Ivshina, M.; Lasko, P.; Richter, J.D. Cytoplasmic polyadenylation element binding proteins in development, health, and disease. *Annu. Rev. Cell Dev. Biol.* **2014**, *30*, 393–415. [CrossRef]
29. Mendez, R.; Richter, J.D. Translational control by CPEB: A means to the end. *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 521–529. [CrossRef]
30. Udagawa, T.; Swanger, S.A.; Takeuchi, K.; Kim, J.H.; Nalavadi, V.; Shin, J.; Lorenz, L.J.; Zukin, R.S.; Bassell, G.J.; Richter, J.D. Bidirectional Control of mRNA Translation and Synaptic Plasticity by the Cytoplasmic Polyadenylation Complex. *Mol. Cell* **2012**, *47*, 253–266. [CrossRef]
31. Lim, J.; Kim, D.; Lee, Y.S.; Ha, M.; Lee, M.; Yeo, J.; Chang, H.; Song, J.; Ahn, K.; Kim, V.N. Mixed tailing by TENT4A and TENT4B shields mRNA from rapid deadenylation. *Science* **2018**, *361*, 701–704. [CrossRef] [PubMed]
32. Lim, J.; Ha, M.; Chang, H.; Kwon, S.C.; Simanshu, D.K.; Patel, D.J.; Kim, V.N. Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell* **2014**, *159*, 1365–1376. [CrossRef]
33. Morgan, M.; Much, C.; DiGiacomo, M.; Azzi, C.; Ivanova, I.; Vitsios, D.M.; Pistolic, J.; Collier, P.; Moreira, P.N.; Benes, V.; et al. MRNA 3′ uridylation and poly(A) tail length sculpt the mammalian maternal transcriptome. *Nature* **2017**, *548*, 347–351. [CrossRef] [PubMed]
34. Kojima, S.; Sher-Chen, E.L.; Green, C.B. Circadian control of mRNA polyadenylation dynamics regulates rhythmic protein expression. *Genes Dev.* **2012**, *26*, 2724–2736. [CrossRef] [PubMed]
35. Liu, Y.; Nie, H.; Lu, F. Dynamic RNA 3′ Uridylation and Guanylation during Mitosis. *iScience* **2020**, *23*. [CrossRef]
36. Park, J.E.; Yi, H.; Kim, Y.; Chang, H.; Kim, V.N. Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. *Mol. Cell* **2016**, *62*, 462–471. [CrossRef] [PubMed]
37. Tian, B.; Manley, J.L. Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.* **2016**, *18*, 18–30. [CrossRef]
38. Sommerkamp, P.; Cabezas-Wallscheid, N.; Trumpp, A. Alternative Polyadenylation in Stem Cell Self-Renewal and Differentiation. *Trends Mol. Med.* **2021**, *27*, 660–672. [CrossRef]
39. Elkon, R.; Ugalde, A.P.; Agami, R. Alternative cleavage and polyadenylation: Extent, regulation and function. *Nat. Rev. Genet.* **2013**, *14*, 496–506. [CrossRef]
40. Lutz, C.S.; Moreira, A. Alternative mRNA polyadenylation in eukaryotes: An effective regulator of gene expression. *Wiley Interdiscip. Rev. RNA* **2011**, *2*, 22–31. [CrossRef]
41. Agarwal, V.; Lopez-Darwin, S.; Kelley, D.R.; Shendure, J. The landscape of alternative polyadenylation in single cells of the developing mouse embryo. *Nat. Commun.* **2021**, *12*, 5101. [CrossRef]
42. Berkovits, B.D.; Mayr, C. Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. *Nature* **2015**, *522*, 363–367. [CrossRef] [PubMed]
43. Jenal, M.; Elkon, R.; Loayza-Puch, F.; Van Haaften, G.; Kühn, U.; Menzies, F.M.; Vrielink, J.A.F.O.; Bos, A.J.; Drost, J.; Rooijers, K.; et al. The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* **2012**, *149*, 538–553. [CrossRef]
44. Batra, R.; Manchanda, M.; Swanson, M.S. Global insights into alternative polyadenylation regulation. *RNA Biol.* **2015**, *12*, 597–602. [CrossRef]
45. Dai, W.; Zhang, G.; Makeyev, E.V. RNA-binding protein HuR autoregulates its expression by promoting alternative polyadenylation site usage. *Nucleic Acids Res.* **2012**, *40*, 787–800. [CrossRef] [PubMed]
46. Batra, R.; Charizanis, K.; Manchanda, M.; Mohan, A.; Li, M.; Finn, D.J.; Goodwin, M.; Zhang, C.; Sobczak, K.; Thornton, C.A.; et al. Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. *Mol. Cell* **2014**, *56*, 311–322. [CrossRef] [PubMed]
47. Sippel, A.E.; Stavrianopoulos, J.G.; Schutz, G.; Feigelson, P. Translational properties of rabbit globin mRNA after specific removal of poly(A) with ribonuclease H. *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 4635–4639. [CrossRef]
48. Sallés, F.J.; Richards, W.G.; Strickland, S. Assaying the polyadenylation state of mRNAs. *Methods* **1999**, *17*, 38–45. [CrossRef]
49. Jänicke, A.; Vancuylenberg, J.; Boag, P.R.; Traven, A.; Beilharz, T.H. ePAT: A simple method to tag adenylated RNA to measure poly(A)-tail length and other 3′ RACE applications. *RNA* **2012**, *18*, 1289–1295. [CrossRef]
50. Murray, E.L.; Schoenberg, D.R. Chapter 24 Assays for Determining Poly(A) Tail Length and the Polarity of mRNA Decay in Mammalian Cells. *Methods Enzymol.* **2008**, *448*, 483–504.
51. Kusov, Y.Y.; Shatirishvili, G.; Dzagurov, G.; Gauss-Müller, V. A new G-tailing method for the determination of the poly(A) tail length applied to hepatitis A virus RNA. *Nucleic Acids Res.* **2001**, *29*, e57. [CrossRef] [PubMed]
52. Garalde, D.R.; Snell, E.A.; Jachimowicz, D.; Sipos, B.; Lloyd, J.H.; Bruce, M.; Pantic, N.; Admassu, T.; James, P.; Warland, A.; et al. Highly parallel direct RN A sequencing on an array of nanopores. *Nat. Methods* **2018**, *15*, 201–206. [CrossRef]
53. Chang, H.; Lim, J.; Ha, M.; Kim, V.N. TAIL-seq: Genome-wide determination of poly(A) tail length and 3′ end modifications. *Mol. Cell* **2014**, *53*, 1044–1052. [CrossRef] [PubMed]
54. Lim, J.; Lee, M.; Son, A.; Chang, H.; Kim, V.N. MTAIL-seq reveals dynamic poly(A) tail regulation in oocyte-to-embryo development. *Genes Dev.* **2016**, *30*, 1671–1682. [CrossRef] [PubMed]

55. Legnini, I.; Alles, J.; Karaiskos, N.; Ayoub, S.; Rajewsky, N. FLAM-seq: Full-length mRNA sequencing reveals principles of poly(A) tail length control. *Nat. Methods* **2019**, *16*, 879–886. [CrossRef] [PubMed]
56. Tushev, G.; Glock, C.; Heumüller, M.; Biever, A.; Jovanovic, M.; Schuman, E.M. Alternative 3′ UTRs Modify the Localization, Regulatory Potential, Stability, and Plasticity of mRNAs in Neuronal Compartments. *Neuron* **2018**, *98*, 495–511.e6. [CrossRef] [PubMed]
57. Ye, C.; Long, Y.; Ji, G.; Li, Q.Q.; Wu, X. APAtrap: Identification and quantification of alternative polyadenylation sites from RNA-seq data. *Bioinformatics* **2018**, *34*, 1841–1849. [CrossRef]
58. Xia, Z.; Donehower, L.A.; Cooper, T.A.; Neilson, J.R.; Wheeler, D.A.; Wagner, E.J.; Li, W. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′-UTR landscape across seven tumour types. *Nat. Commun.* **2014**, *5*, 5274. [CrossRef]
59. Ye, C.; Zhou, Q.; Wu, X.; Yu, C.; Ji, G.; Saban, D.R.; Li, Q.Q. ScDAPA: Detection and visualization of dynamic alternative polyadenylation from single cell RNA-seq data. *Bioinformatics* **2020**, *36*, 1262–1264. [CrossRef]
60. Zhang, Y.; Liu, L.; Qiu, Q.; Zhou, Q.; Ding, J.; Lu, Y.; Liu, P. Alternative polyadenylation: Methods, mechanism, function, and role in cancer. *J. Exp. Clin. Cancer Res.* **2021**, *40*, 51. [CrossRef]
61. Herrmann, C.J.; Schmidt, R.; Kanitz, A.; Artimo, P.; Gruber, A.J.; Zavolan, M. PolyASite 2.0: A consolidated atlas of polyadenylation sites from 3′ end sequencing. *Nucleic Acids Res.* **2020**, *48*, D174–D179. [CrossRef] [PubMed]
62. Müller, S.; Rycak, L.; Afonso-Grunz, F.; Winter, P.; Zawada, A.M.; Damrath, E.; Scheider, J.; Schmäh, J.; Koch, I.; Kahl, G.; et al. APADB: A database for alternative polyadenylation and microRNA regulation events. *Database* **2014**, *2014*, 1–11. [CrossRef] [PubMed]
63. Wang, W.; Wei, Z.; Li, H. A change-point model for identifying 3′UTR switching by next-generation RNA sequencing. *Bioinformatics* **2014**, *30*, 2162–2170. [CrossRef] [PubMed]
64. Grassi, E.; Mariella, E.; Lembo, A.; Molineris, I.; Provero, P. Roar: Detecting alternative polyadenylation with standard mRNA sequencing libraries. *BMC Bioinform.* **2016**, *17*, 423. [CrossRef] [PubMed]
65. Ha, K.C.H.; Blencowe, B.J.; Morris, Q. QAPA: A new method for the systematic analysis of alternative polyadenylation from RNA-seq data. *Genome Biol.* **2018**, *19*, 45. [CrossRef]
66. Arefeen, A.; Liu, J.; Xiao, X.; Jiang, T. TAPAS: Tool for alternative polyadenylation site analysis. *Bioinformatics* **2018**, *34*, 2521–2529. [CrossRef]
67. Gruber, A.J.; Schmidt, R.; Ghosh, S.; Martin, G.; Gruber, A.R.; van Nimwegen, E.; Zavolan, M. Discovery of physiological and cancer-related regulators of 3′ UTR processing with KAPAC. *Genome Biol.* **2018**, *19*, 44. [CrossRef]
68. You, L.; Wu, J.; Feng, Y.; Fu, Y.; Guo, Y.; Long, L.; Zhang, H.; Luan, Y.; Tian, P.; Chen, L.; et al. APASdb: A database describing alternative poly(A) sites and selection of heterogeneous cleavage sites downstream of poly(A) signals. *Nucleic Acids Res.* **2015**, *43*, D59–D67. [CrossRef]
69. Wang, R.; Nambiar, R.; Zheng, D.; Tian, B. PolyA-DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Res.* **2018**, *46*, D315–D319. [CrossRef]
70. Hong, W.; Ruan, H.; Zhang, Z.; Ye, Y.; Liu, Y.; Li, S.; Jing, Y.; Zhang, H.; Diao, L.; Liang, H.; et al. APAatlas: Decoding alternative polyadenylation across human tissues. *Nucleic Acids Res.* **2020**, *48*, D34–D39. [CrossRef]
71. Jin, W.; Zhu, Q.; Yang, Y.; Yang, W.; Wang, D.; Yang, J.; Niu, X.; Yu, D.; Gong, J. Animal-APAdb: A comprehensive animal alternative polyadenylation database. *Nucleic Acids Res.* **2020**, *49*, 47–54. [CrossRef] [PubMed]
72. Malik, I.; Kelley, C.P.; Wang, E.T.; Todd, P.K. Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 589–607. [CrossRef] [PubMed]
73. Ishiura, H.; Tsuji, S. Advances in repeat expansion diseases and a new concept of repeat motif–phenotype correlation. *Curr. Opin. Genet. Dev.* **2020**, *65*, 176–185. [CrossRef]
74. Paulson, H. Repeat Expansion Diseases. *Handb. Clin. Neurol.* **2018**, *147*, 105–123. [PubMed]
75. Krzyzosiak, W.J.; Sobczak, K.; Wojciechowska, M.; Fiszer, A.; Mykowska, A.; Kozlowski, P. Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Res.* **2012**, *40*, 11–26. [CrossRef] [PubMed]
76. Pan, F.; Zhang, Y.; Xu, P.; Man, V.H.; Roland, C.; Weninger, K.; Sagui, C. Molecular conformations and dynamics of nucleotide repeats associated with neurodegenerative diseases: Double helices and CAG hairpin loops. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 2819–2832. [CrossRef]
77. Jae, E.K.; Drier, E.; Barbee, S.A.; Ramaswami, M.; Yin, J.C.P.; Wickens, M. GLD2 poly(A) polymerase is required for long-term memory. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 14644–14649. [CrossRef]
78. Wheeler, J.M.; McMillan, P.; Strovas, T.J.; Liachko, N.F.; Amlie-Wolf, A.; Kow, R.L.; Klein, R.L.; Szot, P.; Robinson, L.; Guthrie, C.; et al. Activity of the poly(A) binding protein MSUT2 determines susceptibility to pathological tau in the mammalian brain. *Sci. Transl. Med.* **2019**, *11*, aao6545. [CrossRef]
79. Morris, K.J.; Corbett, A.H. The polyadenosine RNA-binding protein ZC3H14 interacts with the THO complex and coordinately regulates the processing of neuronal transcripts. *Nucleic Acids Res.* **2018**, *46*, 6561–6575. [CrossRef]
80. Lee, J.; Kim, M.; Itoh, T.Q.; Lim, C. Ataxin-2: A versatile posttranscriptional regulator and its implication in neural function. *Wiley Interdiscip. Rev. RNA* **2018**, *9*, 1–13. [CrossRef]
81. Lastres-Becker, I.; Rüb, U.; Auburger, G. Spinocerebellar ataxia 2 (SCA2). *Cerebellum* **2008**, *7*, 115–124. [CrossRef] [PubMed]
82. Becker, L.A.; Huang, B.; Bieri, G.; Ma, R.; Knowles, D.A.; Jafar-Nejad, P.; Messing, J.; Kim, H.J.; Soriano, A.; Auburger, G.; et al. Therapeutic reduction of ataxin-2 extends lifespan and reduces pathology in TDP-43 mice. *Nature* **2017**, *544*, 367–371. [CrossRef]

83. Tian, B.; Hu, J.; Zhang, H.; Lutz, C.S. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* **2005**, *33*, 201–212. [CrossRef]
84. Lianoglou, S.; Garg, V.; Yang, J.L.; Leslie, C.S.; Mayr, C. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev.* **2013**, *27*, 2380–2396. [CrossRef] [PubMed]
85. Macdonald, C.C. Tissue-specific mechanisms of alternative polyadenylation: Testis, brain, and beyond (2018 update). *RNA* **2019**, *10*, e1526. [CrossRef]
86. Smibert, P.; Miura, P.; Westholm, J.O.; Shenker, S.; May, G.; Duff, M.O.; Zhang, D.; Eads, B.D.; Carlson, J.; Brown, J.B.; et al. Global Patterns of Tissue-Specific Alternative Polyadenylation in Drosophila. *Cell Rep.* **2012**, *1*, 277–289. [CrossRef]
87. Miura, P.; Shenker, S.; Andreu-Agullo, C.; Westholm, J.O.; Lai, E.C. Widespread and extensive lengthening of 39 UTRs in the mammalian brain. *Genome Res.* **2013**, *23*, 812–825. [CrossRef]
88. Wei, L.; Lee, S.; Majumdar, S.; Zhang, B.; Sanfilippo, P.; Joseph, B.; Miura, P.; Soller, M.; Lai, E.C. Overlapping Activities of ELAV/Hu Family RNA Binding Proteins Specify the Extended Neuronal 3′ UTR Landscape in Drosophila. *Mol. Cell* **2020**, *80*, 140–155.e6. [CrossRef]
89. Soller, M.; White, K. ELAV inhibits 3′-end processing to promote neural splicing of ewg pre-mRNA. *Genes Dev.* **2003**, *17*, 2526–2538. [CrossRef] [PubMed]
90. Bae, B.; Miura, P. Emerging roles for 3′ UTRs in neurons. *Int. J. Mol. Sci.* **2020**, *21*, 3413. [CrossRef]
91. Hinman, M.N.; Lou, H. Diverse molecular functions of Hu proteins. *Cell. Mol. Life Sci.* **2008**, *65*, 3168–3181. [CrossRef]
92. Fernandopulle, M.S.; Lippincott-Schwartz, J.; Ward, M.E. RNA transport and local translation in neurodevelopmental and neurodegenerative disease. *Nat. Neurosci.* **2021**, *24*, 622–632. [CrossRef] [PubMed]
93. Blazie, S.M.; Babb, C.; Wilky, H.; Rawls, A.; Park, J.G.; Mangone, M. Comparative RNA-Seq analysis reveals pervasive tissue-specific alternative polyadenylation in Caenorhabditis elegans intestine and muscles. *BMC Biol.* **2015**, *13*, 4. [CrossRef] [PubMed]
94. Relaix, F.; Rocancourt, D.; Mansouri, A.; Buckingham, M. Divergent functions of murine Pax3 and Pax7 in limb muscle development. *Genes Dev.* **2004**, *18*, 1088–1105. [CrossRef]
95. Der Vartanian, A.; Quétin, M.; Michineau, S.; Auradé, F.; Hayashi, S.; Dubois, C.; Rocancourt, D.; Drayton-Libotte, B.; Szegedi, A.; Buckingham, M.; et al. PAX3 Confers Functional Heterogeneity in Skeletal Muscle Stem Cell Responses to Environmental Stress. *Cell Stem Cell* **2019**, *24*, 958–973.e9. [CrossRef] [PubMed]
96. Scaramozza, A.; Park, D.; Kollu, S.; Beerman, I.; Sun, X.; Rossi, D.J.; Lin, C.P.; Scadden, D.T.; Crist, C.; Brack, A.S. Lineage Tracing Reveals a Subset of Reserve Muscle Stem Cells Capable of Clonal Expansion under Stress. *Cell Stem Cell* **2019**, *24*, 944–957.e5. [CrossRef] [PubMed]
97. Boutet, S.C.; Cheung, T.H.; Quach, N.L.; Liu, L.; Prescott, S.L.; Edalati, A.; Iori, K.; Rando, T.A. Alternative polyadenylation mediates microRNA regulation of muscle stem cell function. *Cell Stem Cell* **2012**, *10*, 327–336. [CrossRef] [PubMed]
98. Hirai, H.; Verma, M.; Watanabe, S.; Tastad, C.; Asakura, Y.; Asakura, A. MyoD regulates apoptosis of myoblasts through microRNA-mediated down-regulation of Pax3. *J. Cell Biol.* **2010**, *191*, 347–365. [CrossRef]
99. Crist, C.G.; Montarras, D.; Pallafacchina, G.; Rocancourt, D.; Cumano, A.; Conway, S.J.; Buckinghama, M. Muscle stem cell behavior is modified by microRNA-27 regulation of Pax3 expression. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13383–13387. [CrossRef]
100. Tomé, F.M.S.; Chateau, D.; Helbling-Leclerc, A.; Fardeau, M. Morphological changes in muscle fibers in oculopharyngeal muscular dystrophy. *Neuromuscul. Disord.* **1997**, *7*, 63–69. [CrossRef]
101. Malerba, A.; Klein, P.; Bachtarzi, H.; Jarmin, S.A.; Cordova, G.; Ferry, A.; Strings, V.; Espinoza, M.P.; Mamchaoui, K.; Blumen, S.C.; et al. PABPN1 gene therapy for oculopharyngeal muscular dystrophy. *Nat. Commun.* **2017**, *8*, 14848. [CrossRef]
102. Apponi, L.H.; Leung, S.W.; Williams, K.R.; Valentini, S.R.; Corbett, A.H.; Pavlath, G.K. Loss of nuclear poly(A)-binding protein 1 causes defects in myogenesis and mRNA biogenesis. *Hum. Mol. Genet.* **2010**, *19*, 1058–1065. [CrossRef] [PubMed]
103. Simonelig, M. PABPN1 shuts down alternative poly(A) sites. *Cell Res.* **2012**, *22*, 1419–1421. [CrossRef]
104. Jacquemont, S.; Hagerman, R.J.; Hagerman, P.J.; Leehey, M.A. Fragile-X syndrome and fragile X-associated tremor/ataxia syndrome: Two faces of FMR1. *Lancet Neurol.* **2007**, *6*, 45–55. [CrossRef]
105. Sunamura, N.; Iwashita, S.; Enomoto, K.; Kadoshima, T.; Isono, F. Loss of the fragile X mental retardation protein causes aberrant differentiation in human neural progenitor cells. *Sci. Rep.* **2018**, *8*, 11585. [CrossRef] [PubMed]
106. Cabal-Herrera, A.M.; Tassanakijpanich, N.; Salcedo-Arellano, M.J.; Hagerman, R.J. Fragile X-Associated Tremor/Ataxia Syndrome (FXTAS): Pathophysiology and Clinical Implications. *Int. J. Mol. Sci.* **2020**, *21*, 4391. [CrossRef] [PubMed]
107. Tassone, F.; De Rubeis, S.; Carosi, C.; La Fata, G.; Serpa, G.; Raske, C.; Willemsen, R.; Hagerman, P.J.; Bagni, C. Differential usage of transcriptional start sites and polyadenylation sites in FMR1 premutation alleles. *Nucleic Acids Res.* **2011**, *39*, 6172–6185. [CrossRef]
108. Melamed, Z.; Lopez-Erauskin, J.; Baughn, M.W.; Zhang, O.; Drenner, K.; Sun, Y.; Freyermuth, F.; McMahon, M.A.; Beccari, M.S.; Artates, J.; et al. Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration. *Nat. Neurosci.* **2019**, *22*, 180. [CrossRef]
109. Patel, R.; Brophy, C.; Hickling, M.; Neve, J.; Furger, A. Alternative cleavage and polyadenylation of genes associated with protein turnover and mitochondrial function are deregulated in Parkinson's, Alzheimer's and ALS disease. *BMC Med. Genom.* **2019**, *12*, 1–14. [CrossRef]
110. Balendra, R.; Isaacs, A.M. C9orf72-mediated ALS and FTD: Multiple pathways to disease. *Nat. Rev. Neurol.* **2018**, *14*, 544–558. [CrossRef]

111. Niblock, M.; Smith, B.N.; Lee, Y.B.; Sardone, V.; Topp, S.; Troakes, C.; Al-Sarraj, S.; Leblond, C.S.; Dion, P.A.; Rouleau, G.A.; et al. Retention of hexanucleotide repeat-containing intron in C9orf72 mRNA: Implications for the pathogenesis of ALS/FTD. *Acta Neuropathol. Commun.* **2016**, *4*, 18. [CrossRef]
112. Prudencio, M.; Belzil, V.V.; Batra, R.; Ross, C.A.; Gendron, T.F.; Pregent, L.J.; Murray, M.E.; Overstreet, K.K.; Piazza-Johnston, A.E.; Desaro, P.; et al. Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nat. Neurosci.* **2015**, *18*, 1175–1182. [CrossRef] [PubMed]
113. Giannini, M.; Bayona-Feliu, A.; Sproviero, D.; Barroso, S.I.; Cereda, C.; Aguilera, A. TDP-43 mutations link Amyotrophic Lateral Sclerosis with R-loop homeostasis and R loop-mediated DNA damage. *PLoS Genet.* **2020**, *16*, e1009260. [CrossRef] [PubMed]
114. Prashad, S.; Gopal, P.P. RNA-binding proteins in neurological development and disease. *RNA Biol.* **2021**, *18*, 972. [CrossRef] [PubMed]
115. Humphrey, J.; Birsa, N.; Milioto, C.; McLaughlin, M.; Ule, A.M.; Robaldo, D.; Eberle, A.B.; Kräuchi, R.; Bentham, M.; Brown, A.L.; et al. FUS ALS-causative mutations impair FUS autoregulation and splicing factor networks through intron retention. *Nucleic Acids Res.* **2020**, *48*, 6889–6905. [CrossRef] [PubMed]
116. Koyama, A.; Sugai, A.; Kato, T.; Ishihara, T.; Shiga, A.; Toyoshima, Y.; Koyama, M.; Konno, T.; Hirokawa, S.; Yokoseki, A.; et al. Increased cytoplasmic TARDBP mRNA in affected spinal motor neurons in ALS caused by abnormal autoregulation of TDP-43. *Nucleic Acids Res.* **2016**, *44*, 5820–5836. [CrossRef] [PubMed]
117. LoRusso, S.; Weiner, B.; Arnold, W.D. Myotonic Dystrophies: Targeting Therapies for Multisystem Disease. *Neurotherapeutics* **2018**, *15*, 872–884. [CrossRef]
118. Ozimski, L.L.; Sabater-Arcis, M.; Bargiela, A.; Artero, R. The hallmarks of myotonic dystrophy type 1 muscle dysfunction. *Biol. Rev.* **2021**, *96*, 716–730. [CrossRef]
119. Kurkiewicz, A.; Cooper, A.; McIlwaine, E.; Cumming, S.A.; Adam, B.; Krahe, R.; Puymirat, J.; Schoser, B.; Timchenko, L.; Ashizawa, T.; et al. Towards development of a statistical framework to evaluate myotonic dystrophy type 1 mRNA biomarkers in the context of a clinical trial. *PLoS ONE* **2020**, *15*, e0231000. [CrossRef]
120. Gudde, A.E.E.G.; van Kessel, I.D.G.; André, L.M.; Wieringa, B.; Wansink, D.G. Trinucleotide-repeat expanded and normal DMPK transcripts contain unusually long poly(A) tails despite differential nuclear residence. *Biochim. Biophys. Acta—Gene Regul. Mech.* **2017**, *1860*, 740–749. [CrossRef]
121. Kino, Y.; Mori, D.; Oma, Y.; Takeshita, Y.; Sasagawa, N.; Ishiura, S. Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum. Mol. Genet.* **2004**, *13*, 495–507. [CrossRef]
122. Yuan, Y.; Compton, S.A.; Sobczak, K.; Stenberg, M.G.; Thornton, C.A.; Griffith, J.D.; Swanson, M.S. Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res.* **2007**, *35*, 5474–5486. [CrossRef] [PubMed]
123. Konieczny, P.; Stepniak-Konieczna, E.; Sobczak, K. MBNL proteins and their target RNAs, interaction and splicing regulation. *Nucleic Acids Res.* **2014**, *42*, 10873–10887. [CrossRef] [PubMed]
124. Taylor, K.; Sznajder, Ł.J.; Cywoniuk, P.; Thomas, J.D.; Swanson, M.S.; Sobczak, K. MBNL splicing activity depends on RNA binding site structural context. *Nucleic Acids Res.* **2018**, *46*, 9119–9133. [CrossRef]
125. Brinegar, A.E.; Cooper, T.A. Roles for RNA-binding proteins in development and disease. *Brain Res.* **2016**, *1647*, 1–8. [CrossRef] [PubMed]
126. Wang, E.T.; Cody, N.A.L.; Jog, S.; Biancolella, M.; Wang, T.T.; Treacy, D.J.; Luo, S.; Schroth, G.P.; Housman, D.E.; Reddy, S.; et al. Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* **2012**, *150*, 710–724. [CrossRef] [PubMed]
127. Itskovich, S.S.; Gurunathan, A.; Clark, J.; Burwinkel, M.; Wunderlich, M.; Berger, M.R.; Kulkarni, A.; Chetal, K.; Venkatasubramanian, M.; Salomonis, N.; et al. MBNL1 regulates essential alternative RNA splicing patterns in MLL-rearranged leukemia. *Nat. Commun.* **2020**, *11*, 2369. [CrossRef]
128. Mankodi, A. Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum. Mol. Genet.* **2001**, *10*, 2165–2170. [CrossRef]
129. Jiang, H.; Mankodi, A.; Swanson, M.S.; Moxley, R.T.; Thornton, C.A. Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum. Mol. Genet.* **2004**, *13*, 3079–3088. [CrossRef]
130. Song, K.Y.; Guo, X.M.; Wang, H.Q.; Zhang, L.; Huang, S.Y.; Huo, Y.C.; Zhang, G.; Feng, J.Z.; Zhang, R.R.; Ma, Y.; et al. MBNL1 reverses the proliferation defect of skeletal muscle satellite cells in myotonic dystrophy type 1 by inhibiting autophagy via the mTOR pathway. *Cell Death Dis.* **2020**, *11*, 545. [CrossRef]
131. 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. *Cell* **1993**, *72*, 971–983. [CrossRef]
132. Tabrizi, S.J.; Flower, M.D.; Ross, C.A.; Wild, E.J. Huntington disease: New insights into molecular pathogenesis and therapeutic opportunities. *Nat. Rev. Neurol.* **2020**, *16*, 529–546. [CrossRef]
133. Mccolgan, P.; Tabrizi, S.J. Huntington's disease: A clinical review. *Eur. J. Neurol.* **2018**, *25*, 24–34. [CrossRef] [PubMed]
134. Romo, L.; Ashar-Patel, A.; Pfister, E.; Aronin, N. Alterations in mRNA 3′ UTR Isoform Abundance Accompany Gene Expression Changes in Human Huntington's Disease Brains. *Cell Rep.* **2017**, *20*, 3057–3070. [CrossRef]

135. Lin, B.; Rommens, J.M.; Graham, R.K.; Kalchman, M.; Macdonald, H.; Nasir, J.; Delaney, A.; Goldberg, Y.P.; Hayden, M.R. Differential 3′ polyadenylation of the huntington disease gene results in two mRNA species with variable tissue expression. *Hum. Mol. Genet.* **1993**, *2*, 1541–1545. [CrossRef] [PubMed]
136. Xu, H.; An, J.J.; Xu, B. Distinct cellular toxicity of two mutant huntingtin mRNA variants due to translation regulation. *PLoS ONE* **2017**, *12*, e0177610. [CrossRef] [PubMed]
137. Sathasivam, K.; Neueder, A.; Gipson, T.A.; Landles, C.; Benjamin, A.C.; Bondulich, M.K.; Smith, D.L.; Faull, R.L.M.; 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]
138. Neueder, A.; Landles, C.; Ghosh, R.; Howland, D.; Myers, R.H.; Faull, R.L.M.; Tabrizi, S.J.; Bates, G.P. The pathogenic exon 1 HTT protein is produced by incomplete splicing in Huntington's disease patients. *Sci. Rep.* **2017**, *7*, 1307. [CrossRef]
139. Neueder, A.; Dumas, A.A.; Benjamin, A.C.; Bates, G.P. Regulatory mechanisms of incomplete huntingtin mRNA splicing. *Nat. Commun.* **2018**, *9*, 3955. [CrossRef]
140. Mason, M.A.; Gomez-Paredes, C.; Sathasivam, K.; Neueder, A.; Papadopoulou, A.S.; Bates, G.P. Silencing Srsf6 does not modulate incomplete splicing of the huntingtin gene in Huntington's disease models. *Sci. Rep.* **2020**, *10*, 14057. [CrossRef]
141. Lieberman, A.P.; Shakkottai, V.G.; Albin, R.L. Polyglutamine Repeats in Neurodegenerative Diseases. *Annu. Rev. Pathol. Mech. Dis.* **2019**, *14*, 1–27. [CrossRef]
142. Kamieniarz-Gdula, K.; Proudfoot, N.J. Transcriptional Control by Premature Termination: A Forgotten Mechanism. *Trends Genet.* **2019**, *35*, 553–564. [CrossRef]
143. Ciesiolka, A.; Jazurek, M.; Drazkowska, K.; Krzyzosiak, W.J. Structural characteristics of simple RNA repeats associated with disease and their deleterious protein interactions. *Front. Cell. Neurosci.* **2017**, *11*, 1–19. [CrossRef]
144. Howe, K.L.; Achuthan, P.; Allen, J.; Allen, J.; Alvarez-Jarreta, J.; Ridwan Amode, M.; Armean, I.M.; Azov, A.G.; Bennett, R.; Bhai, J.; et al. Ensembl 2021. *Nucleic Acids Res.* **2021**, *49*, D884–D891. [CrossRef]
145. Chen, C.Y.; Chen, S.T.; Juan, H.F.; Huang, H.C. Lengthening of 3′UTR increases with morphological complexity in animal evolution. *Bioinformatics* **2012**, *28*, 3178–3181. [CrossRef]
146. Mayr, C. What are 3′ utrs doing? *Cold Spring Harb. Perspect. Biol.* **2019**, *11*. [CrossRef]
147. Zhu, S.; Lian, Q.; Ye, W.; Qin, W.; Wu, Z.; Ji, G.; Wu, X. scAPAdb: A comprehensive database of alternative polyadenylation at single-cell resolution. *Nucleic Acids Res.* **2021**, gkab795. [CrossRef]
148. Gao, Y.; Li, L.; Amos, C.I.; Li, W. Analysis of alternative polyadenylation from singlecell RNA-seq using scDaPars reveals cell subpopulations invisible to gene expression. *Genome Res.* **2021**, *31*, 1856–1866. [CrossRef]
149. Saxena, S.; Caroni, P. Selective Neuronal Vulnerability in Neurodegenerative Diseases: From Stressor Thresholds to Degeneration. *Neuron* **2011**, *71*, 35–48. [CrossRef]
150. Ogorodnikov, A.; Levin, M.; Tattikota, S.; Tokalov, S.; Hoque, M.; Scherzinger, D.; Marini, F.; Poetsch, A.; Binder, H.; Macher-Göppinger, S.; et al. Transcriptome 3′end organization by PCF11 links alternative polyadenylation to formation and neuronal differentiation of neuroblastoma. *Nat. Commun.* **2018**, *9*, 5331. [CrossRef]
151. Masamha, C.P.; Xia, Z.; Yang, J.; Albrecht, T.R.; Li, M.; Shyu, A.-B.; Li, W.; Wagner, E.J. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature* **2014**, *510*, 412–416. [CrossRef]
152. Fiszer, A.; Krzyzosiak, W.J. Oligonucleotide-based strategies to combat polyglutamine diseases. *Nucleic Acids Res.* **2014**, *42*, 6787–6810. [CrossRef] [PubMed]
153. 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]
154. Ciesiolka, A.; Stroynowska-Czerwinska, A.; Joachimiak, P.; Ciolak, A.; Kozlowska, E.; Michalak, M.; Dabrowska, M.; Olejniczak, M.; Raczynska, K.D.; Zielinska, D.; et al. Artificial miRNAs targeting CAG repeat expansion in ORFs cause rapid deadenylation and translation inhibition of mutant transcripts. *Cell. Mol. Life Sci.* **2020**. [CrossRef]

Table S1. Repeat expansion diseases.

	Disease	Mutant gene	Repeat motif	Repeat tract location
1	**Benign adult familial myoclonic epilepsy (BAFME)**	*SAMD12*	e.g. TTTCA	Intron
2	**Baratela-Scott Syndrome**	*XYLT1*	GGC	Promoter
3	**C9ORF72-mediated ALS/FTD**	*C9ORF72*	GGGGCC	Intron
4	**Cerebellar ataxia, neuropathy, and vestibular areflexia syndrome (CANVAS)**	*RFC1*	AAGGG	Intron
5	**Cleidocranial dysplasia (CCD)**	*RUNX2*	GCN	Exon
6	**Congenital central hypoventilation syndrome (CCHS)**	*PHOX2B*	GCN	Exon
7	**Myotonic dystrophy type 1 (DM1)**	*DMPK*	CTG	3'-UTR
8	**Myotonic dystrophy type 2 (DM2)**	*CNBP1*	CCTG	Intron
9	**Dentatorubro-pallidoluysian atrophy (DRPLA)**	*ATN1*	CAG	Exon
10	**Familial adult myoclonic epilepsy types 1-7 (FAME 1-7)**	Various	TTTCA	Intron
11	**Fuchs endothelial corneal dystrophy (FECD)**	*TCF4*	CTG	Intron
12	**Fragile XE syndrome (FRAXE)**	*AFF2/FMR2*	CCG	5'-UTR
13	**Friedreich ataxia (FRDA)**	*FXN*	GAA	Intron
14	**Fragile X syndrome; Fragile X-associated tremor/ataxia syndrome (FXS; FXTAS)**	*FMR1*	CGG	5'-UTR
15	**Glutaminase deficiency**	*GLS*	CAG	5' UTR
16	**Huntington's disease (HD)**	*HTT*	CAG	Exon
17	**Huntington disease-like 2 (HDL2)**	*JPH3*	CTG	Exon
18	**Hand–foot–genital syndrome (HFGS)**	*HOXA13*	GCN	Exon
19	**Holoprosencephaly 5 (HPE5)**	*ZIC2*	GCN	Exon
20	**Jacobsen syndrome**	*CBL2*	CCG	5'-UTR
21	**Neuronal intranuclear inclusion disease (NIID)**	*NOTCH2/NLC*	CGG	5'-UTR

22	**Oculopharyngeal muscular dystrophy (OPMD)**	*PAPBN1*	GCG	Exon
23	**X-linked hypopituitarism (PHPX)**	*SOX3*	GCN	Exon
24	**Spinal and bulbar muscular atrophy (SBMA)**	*AR*	CAG	Exon
25	**Spinocerebellar ataxia type 1 (SCA1)**	*ATXN1*	CAG	Exon
26	**Spinocerebellar ataxia type 2 (SCA2)**	*ATXN2*	CAG	Exon
27	**Spinocerebellar ataxia type 3 (SCA3)**	*ATXN3*	CAG	Exon
28	**Spinocerebellar ataxia type 6 (SCA6)**	*CACNA1A*	CAG	Exon
29	**Spinocerebellar ataxia type 7 (SCA7)**	*ATXN7*	CAG	Exon
30	**Spinocerebellar ataxia type 8 (SCA8)**	*ATXN8/ATXN8OS*	CAG/CTG	3'-UTR
31	**Spinocerebellar ataxia type 10 (SCA10)**	*ATXN10*	ATTCT	Intron
32	**Spinocerebellar ataxia type 12 (SCA12)**	*PPP2R2B*	CAG	5'-UTR
33	**Spinocerebellar ataxia type 17 (SCA17)**	*TBP*	CAG	Exon
34	**Spinocerebellar ataxia type 31 (SCA31)**	*TK2* and *BEAN*	TGGA	Intron
35	**Spinocerebellar ataxia type 36 (SCA36)**	*NOP56*	GGCCTG	Intron
36	**Spinocerebellar ataxia type 37 (SCA37)**	*DAB1*	ATTTC	Intron
37	**Unverricht–Lundborg disease (ULD)**	*CSTB*	CCCCGCCCCGCG	Promoter
38	**X-linked dystonia parkinsonism (XDP)**	*TAF1*	CCCTCT	Intron
39	**X-linked intellectual disability (XLMR)**	*ARX*	GCG	Exon

BMC Biology

RESEARCH ARTICLE

Open Access

Allele-specific quantitation of *ATXN3* and *HTT* transcripts in polyQ disease models

Paweł Joachimiak[1], Adam Ciesiołka[1], Emilia Kozłowska[1], Paweł M. Świtoński[1], Grzegorz Figura[1], Agata Ciołak[1], Grażyna Adamek[1], Magdalena Surdyka[2], Żaneta Kalinowska-Pośka[2], Maciej Figiel[2], Nicholas S. Caron[3], Michael R. Hayden[3] and Agnieszka Fiszer[1*]

Abstract

Background The majority of genes in the human genome is present in two copies but the expression levels of both alleles is not equal. Allelic imbalance is an aspect of gene expression relevant not only in the context of genetic variation, but also to understand the pathophysiology of genes implicated in genetic disorders, in particular, dominant genetic diseases where patients possess one normal and one mutant allele. Polyglutamine (polyQ) diseases are caused by the expansion of CAG trinucleotide tracts within specific genes. Spinocerebellar ataxia type 3 (SCA3) and Huntington's disease (HD) patients harbor one normal and one mutant allele that differ in the length of CAG tracts. However, assessing the expression level of individual alleles is challenging due to the presence of abundant CAG repeats in the human transcriptome, which make difficult the design of allele-specific methods, as well as of therapeutic strategies to selectively engage CAG sequences in mutant transcripts.

Results To precisely quantify expression in an allele-specific manner, we used SNP variants that are linked to either normal or CAG expanded alleles of the ataxin-3 (*ATXN3*) and huntingtin (*HTT*) genes in selected patient-derived cell lines. We applied a SNP-based quantitative droplet digital PCR (ddPCR) protocol for precise determination of the levels of transcripts in cellular and mouse models. For HD, we showed that the process of cell differentiation can affect the ratio between endogenous alleles of *HTT* mRNA. Additionally, we reported changes in the absolute number of the *ATXN3* and *HTT* transcripts per cell during neuronal differentiation. We also implemented our assay to reliably monitor, in an allele-specific manner, the silencing efficiency of mRNA-targeting therapeutic approaches for HD. Finally, using the humanized Hu128/21 HD mouse model, we showed that the ratio of normal and mutant *HTT* transgene expression in brain slightly changes with the age of mice.

Conclusions Using allele-specific ddPCR assays, we observed differences in allele expression levels in the context of SCA3 and HD. Our allele-selective approach is a reliable and quantitative method to analyze low abundant transcripts and is performed with high accuracy and reproducibility. Therefore, the use of this approach can significantly improve understanding of allele-related mechanisms, e.g., related with mRNA processing that may be affected in polyQ diseases.

Keywords Polyglutamine diseases, Huntington's disease, Spinocerebellar ataxia type 3, ddPCR, Allele-specific quantitation, SNP

*Correspondence:
Agnieszka Fiszer
agnieszka.fiszer@ibch.poznan.pl
Full list of author information is available at the end of the article

Background

Polyglutamine (polyQ) diseases are a family of neurodegenerative disorders that include spinocerebellar ataxia type 3 (SCA3) and Huntington's disease (HD), caused by a CAG repeat expansion in the coding region of specific genes [1, 2]. Due to an autosomal dominant inheritance pattern, most of polyQ disease patients carry a normal (wild type, WT) and mutant (MUT) allele. Normal alleles usually contain 5–30 CAG repeats, while mutant alleles, encoding proteins with long stretches of glutamines, are characterized by more than 39 CAG repeats in HD and above 60 CAG repeats in the case of SCA3 [2, 3].

SCA3 is caused by a CAG repeat expansion in the exon 10 of the ataxin-3 (*ATXN3*) gene and typically manifests with dysfunction and degeneration of neurons in cerebellum and spinocerebellar tracts [4]. HD, in turn, is caused by a CAG repeat expansion in the first exon of the huntingtin (*HTT*) gene. In HD, the mutation leads mostly to degeneration of striatal and cortical neurons. However, other brain regions may also be substantially affected in both SCA3 and HD [5].

A mutant protein is generally recognized to be the main pathogenic factor in polyQ diseases. Expanded polyQ tracts induce the formation of misfolded protein aggregates that disturb cellular homeostasis and lead to neuronal death [2, 6–9]. The exact role of mutant transcripts in pathogenic mechanisms in polyQ diseases is being unraveled [10–13]. RNA-related studies have demonstrated so far: incomplete splicing of *HTT* mRNA [14], abnormal interactions of RNAs containing expanded CAG repeats with proteins [15, 16], and RNA foci formation [17, 18], and potential importance of alternative polyadenylation of various polyQ transcripts [19]. Furthermore, mutant polyQ disease transcripts have been shown to be promising therapeutic targets in strategies that aim to downregulate mutant gene expression [2, 9, 20].

Molecular examination of endogenous polyQ transcripts is a challenging task. PolyQ disease genes are expressed at low levels across all tissues, including the brain regions that are mostly affected by the disease [21]. Any method that seeks to inspect polyQ disease transcripts needs to be highly accurate and sensitive. Moreover, if method of choice requires discrimination between WT and MUT alleles, the CAG sequence region cannot be used to design a primer or probe. One possible solution to overcome these limitations is to utilize heterozygous SNP variants present in transcripts. Droplet digital PCR (ddPCR) utilizes droplet-based microfluidics and compartmentalization to provide absolute quantification of analyzed molecules and has been used for allele-selective quantification of transcripts [22–25].

In this study, we first identified heterozygous SNP variants in *ATXN3* and *HTT* cDNAs from patient-derived cell lines. Next, we adapted ddPCR to perform accurate and quantitative determination of the WT/MUT allele ratio and estimation of the number of the *ATXN3* and *HTT* transcripts per cell. We performed our analyses on fibroblasts and induced pluripotent stem cells (iPSCs), as well as neural stem cells (NSCs) and neurons, differentiated from iPSCs. Using these cell lines, we demonstrated changes in WT/MUT allele expression ratios emerging from a neuronal differentiation process. Additionally, we determined the number of *HTT* WT and MUT transgene copies and their expression ratio in the brain tissue of biallelic HD mice. The obtained results provide a starting point for further investigation of the relevance of expression of WT and MUT alleles to the pathology of polyQ diseases.

Results

Examination of endogenous ATXN3 transcripts

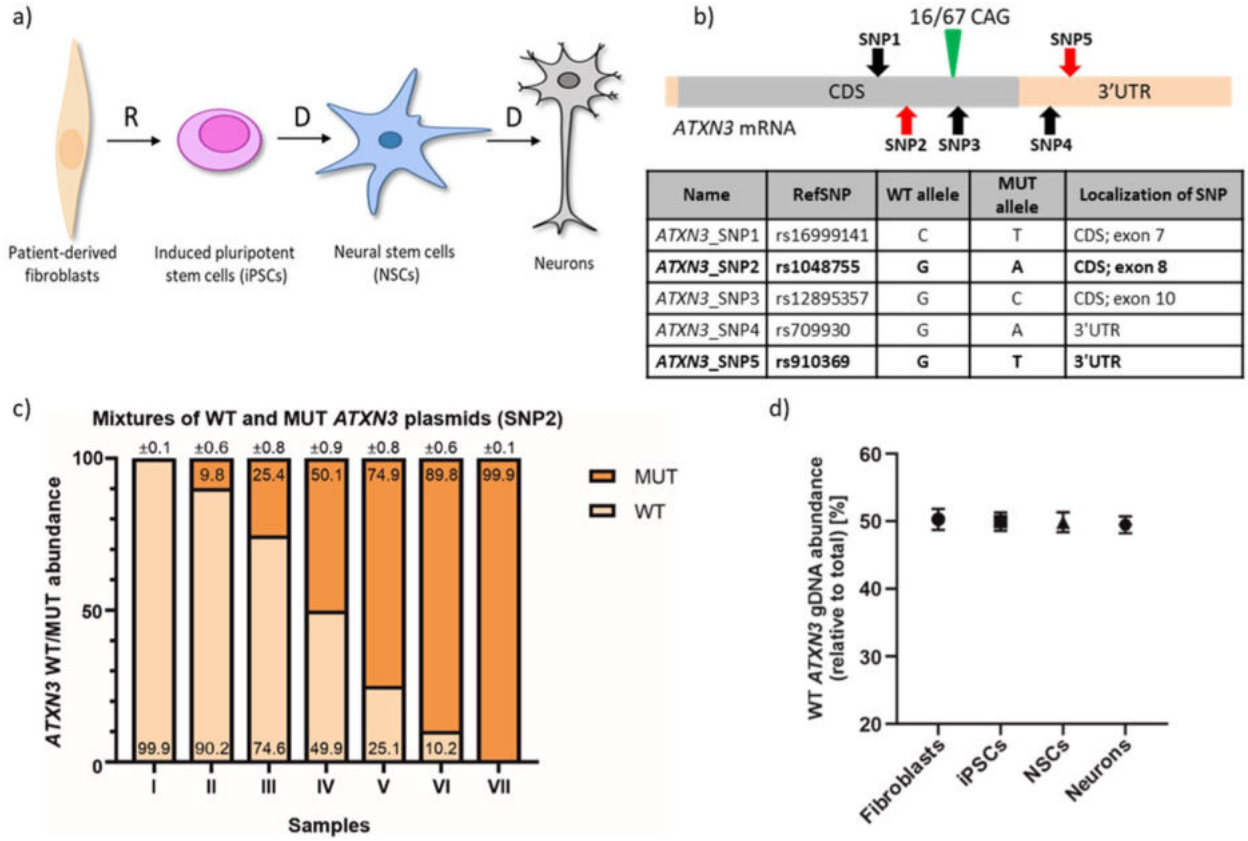
For the ddPCR assay measuring the expression of *ATXN3* WT and MUT alleles, we used a set of patient-derived cell lines with the SCA3 genetic background: fibroblasts, iPSCs, NSCs, and neurons (characterized by 16/67 CAG repeats in *ATXN3*) (Fig. 1a).

Identification of SNP variants in ATXN3 transcripts

To identify SNP variants, we cloned and sequenced *ATXN3* cDNA from SCA3 NSCs and identified 5 heterozygous SNPs in exons of the *ATXN3* gene (Fig. 1b). Two of these *ATXN3* SNPs (rs1048755 and rs910369, referred to as *ATXN3*_SNP2 and *ATXN3*_SNP5) were selected to design ddPCR assays. These SNPs either lead to missense mutation in N-terminal ubiquitin-interacting motif 1 (UIM1) of ataxin-3 [26] or are considered to cause alterations of RNA-binding protein (RBP)-binding sites in *ATXN3* mRNA [27].

Validation of the ATXN3 ddPCR assay specificity and accuracy

To examine specificity of allelic discrimination of *ATXN3* ddPCR assays, we performed *ATXN3*_SNP2 and *ATXN3*_SNP5 ddPCRs on predefined mixtures of plasmids containing either WT or MUT sequences with respective SNPs. We observed that the ratio of the signal originating from WT and MUT probes was extremely consistent with the ratio of premixed WT/MUT plasmids for both SNPs: SNP2 (Fig. 1c) and SNP5 (Additional File 1: Fig. S1A). To further validate the *ATXN3* ddPCR accuracy, we examined gDNA isolated from SCA3 fibroblasts, iPSCs, NSCs, and neurons. We used the *ATXN3*_SNP5 assay that utilizes the SNP variant located in the 3′UTR of the *ATXN3* gene, rendering the assay suitable for gDNA analysis. We found that in all analyzed gDNA



Name	RefSNP	WT allele	MUT allele	Localization of SNP
*ATXN3*_SNP1	rs16999141	C	T	CDS; exon 7
***ATXN3*_SNP2**	**rs1048755**	**G**	**A**	**CDS; exon 8**
*ATXN3*_SNP3	rs12895357	G	C	CDS; exon 10
*ATXN3*_SNP4	rs709930	G	A	3'UTR
***ATXN3*_SNP5**	**rs910369**	**G**	**T**	**3'UTR**

Fig. 1 SNP-based allele-specific ddPCR assays for ATXN3 transcript. ***a*** SCA3 cell lines used; R—reprogramming; D—differentiation. ***b*** Scheme of ATXN3 transcript with marked CAG repeat tract (green triangle) and SNP variants identified in SCA3 cell lines (red arrows—SNPs used in ddPCR assays; black arrows—other identified SNPs); light orange box—UTRs. Table presents all SNP variants identified in analyzed SCA3 cell lines. Bolded SNPs were selected to be targets for ddPCR assays, and they will be hereinafter referred to as ATXN3_SNP2 and ATXN3_SNP5 in the text. CDS—coding sequence. RefSNP number according to the Single Nucleotide Polymorphism Database (dbSNP). ***c*** Results from ddPCR analysis of ATXN3_SNP2 assay specificity performed on seven samples with predefined ratios of WT/MUT ATXN3 plasmids (samples: I—100% WT and 0% MUT; II—90% WT and 10% MUT; III—75% WT and 25% MUT; IV—50% WT and 50% MUT; V—25% WT and 75% MUT; VI—10% WT and 90% MUT; VII—0% WT and 100% MUT). Precise values are indicated on WT/MUT bars ± poisson error. ***d*** Percentage ratio of ATXN3 WT allele obtained with ddPCR ATXN3_SNP5 assay using gDNA from indicated SCA3 cell lines. Three biological replicates were performed. These data are presented as means ± SD

samples, the WT/MUT allele ratio was remarkably close to the expected 50:50 (Fig. 1d).

Determination of the ATXN3 mRNA allele ratio in SCA3 cell lines

We examined allele-specific expression patterns of *ATXN3* in SCA3 cell lines: patient-derived fibroblasts, iPSCs reprogrammed from fibroblasts, as well as NSCs and neurons differentiated from iPSCs (Fig. 1a). To confirm cellular identity of generated neurons, we performed immunostaining and observed strong expression of *MAP 2* and *TUJ1* neuronal markers (Additional File 2: Fig. S2). When we quantified the *ATXN3* allelic expression ratios in SCA3 cell lines, we observed 54.7% WT/45.3% MUT ratio in fibroblasts, 53.2% WT/46.8% MUT ratio in iPSCs, 53.9% WT/46.1% MUT in NSCs, and 53.8% WT/46.2% MUT ratio in neurons (Fig. 2a–d, precise values for each *ATXN3* assay are given in Additional File 3: Fig. S3). We detected significantly higher expression of the WT allele, as compared to MUT allele, in almost all analyzed cell types. Two-way ANOVA showed that there are no statistically significant changes in the WT abundance between cell types (Fig. 2e); therefore, we conclude that the ratio of endogenous *ATXN3* alleles is not influenced by the cell type (in analyzed set of cell lines).

Estimation of the number of endogenous ATXN3 transcripts per diploid genome

To examine whether there are differences in the total number of *ATXN3* transcripts among analyzed cell lines, we split cells into two tubes and performed concurrent RNA and DNA isolations. We used the 3′UTR-located *ATXN3*_SNP5 assay to analyze both cDNA and gDNA. By dividing the total copy number of *ATXN3* transcripts

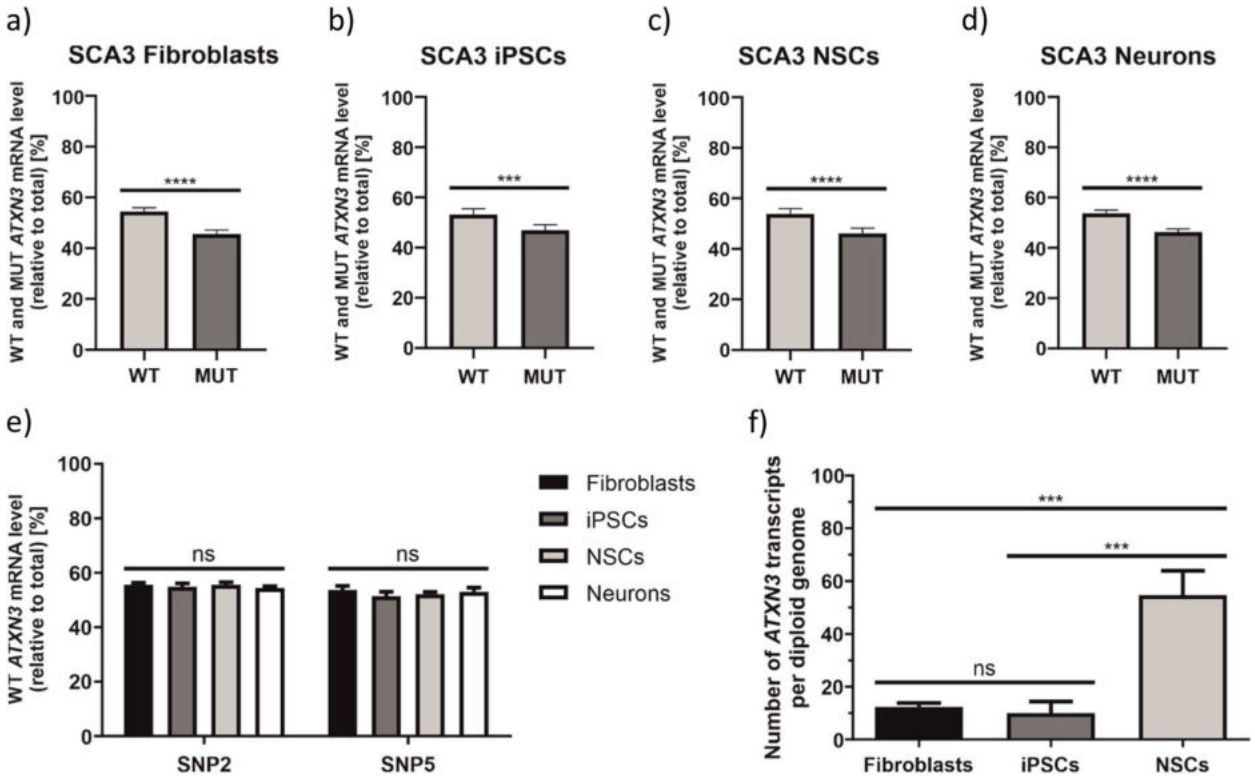


Fig. 2 Determination of endogenous WT and MUT ATXN3 allele expression ratios in SCA3 cell lines. **a–d** Results from ddPCR are presented as a mean WT and MUT ATXN3 transcript allele abundance, calculated based on results from both ATXN3 assays (ATXN3_SNP2 and ATXN3_SNP5), in patient-derived fibroblasts (**a**), iPSCs (**b**), NSCs (**c**), and neurons (**d**). These data were analyzed using unpaired t test. **e** Corresponding results to **a–d** presented for ATXN3_SNP2 and ATXN3_SNP5 assays separately and analyzed with two-way ANOVA with Tukey's multiple comparison test for identification of any cell type-specific changes. In this graph, only WT allele abundance is presented, the MUT allele abundance equals a remainder to the sum of 100%). **f** Estimation of the total number (WT + MUT) of endogenous ATXN3 transcripts per diploid genome using ATXN3_SNP5 assay. Data were analyzed using one-way ANOVA with Tukey's multiple comparison test. For all experiments presented in this figure $n=3$. Two-tailed p value < 0.05 was considered significant and is depicted in the figure by: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. All data are presented as means ± SD. Individual data values are available in Additional File 13

by the total copy number of gDNA amplicons, we estimated a number of *ATXN3* transcripts per diploid genome. We documented approximately 12, 10, and 54 copies of *ATXN3* mRNA in fibroblasts, iPSCs, and NSCs, respectively (Fig. 2f). This result suggests the increase of *ATXN3* expression in the process of neural differentiation. Moreover, the number of *ATXN3* transcripts calculated for fibroblasts corresponds well with our previous results generated using single-molecule FISH, for this cell line (approximately 18 copies of *ATXN3* mRNA per cell) [28].

Examination of endogenous HTT transcripts in cell lines

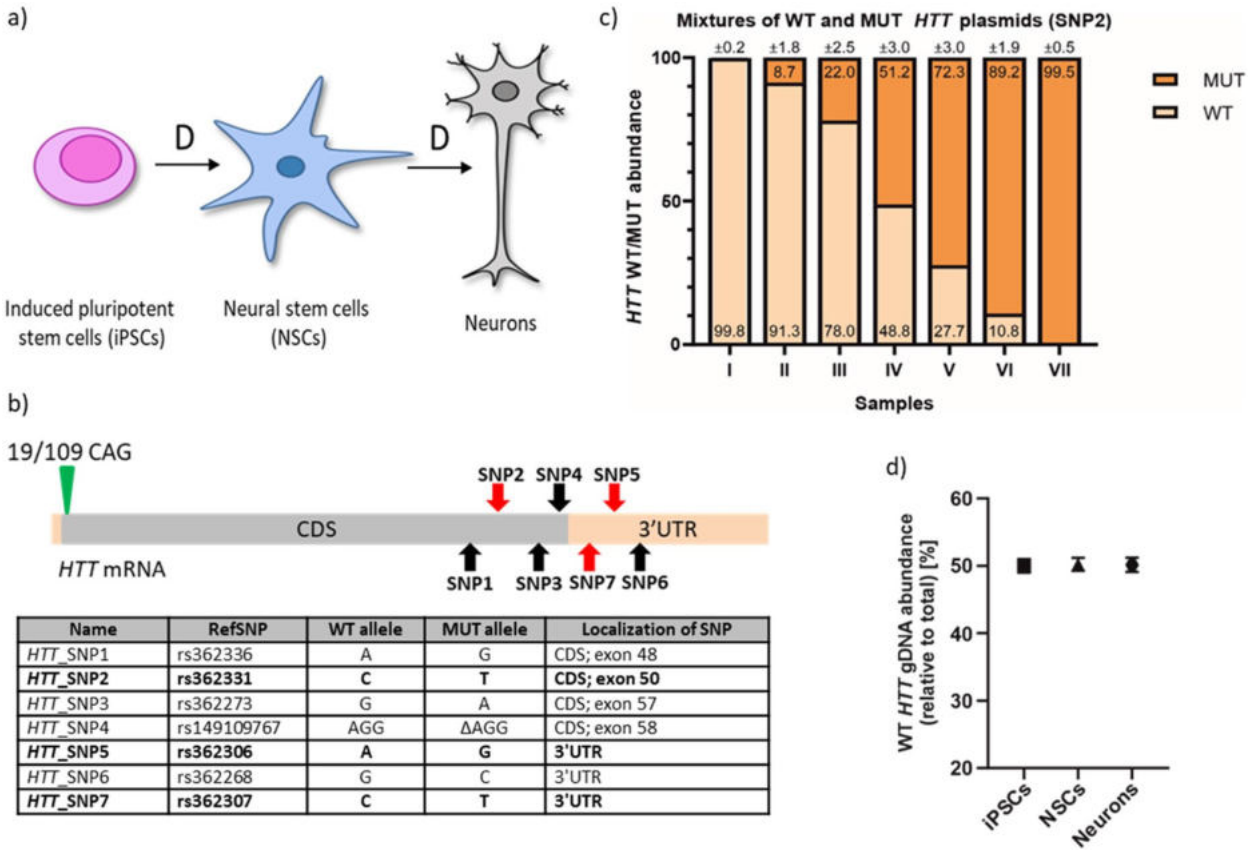
We used a set of patient-derived cell lines with the HD genetic background: iPSCs, NSCs, and neurons (characterized by 19/109 CAG repeats in *HTT*) for ddPCR assay discriminating between expression of *HTT* WT and MUT allele (Fig. 3a).

Identification of SNP variants in HTT transcripts

To identify SNP variants located in *HTT* transcripts, we replicated our approach implemented for *ATXN3*. First, through the cloning and sequencing process, we identified 7 heterozygous SNPs in *HTT* exons in HD patient-derived NSCs (Fig. 3b). Three of them (rs362331, rs362306, and rs362307, named *HTT*_SNP2, *HTT*_SNP5 and *HTT*_SNP7, respectively) were selected for design of ddPCR assays (Fig. 3b). Those SNPs were described as associated with the most common *HTT* haplotype [29], correlated with the presence of mutation in *HTT* allele [30], and/or were targets in clinical testing of therapeutic approaches using ASOs.

Validation of specificity and utility of the HTT ddPCR assays

To assess allelic specificity of the HD ddPCR assay, we prepared plasmid mixtures of predefined WT/MUT ratios. We confirmed that ddPCR output accurately reflects tested plasmid ratios, using all three assays:



Name	RefSNP	WT allele	MUT allele	Localization of SNP
*HTT*_SNP1	rs362336	A	G	CDS; exon 48
***HTT*_SNP2**	**rs362331**	**C**	**T**	**CDS; exon 50**
*HTT*_SNP3	rs362273	G	A	CDS; exon 57
*HTT*_SNP4	rs149109767	AGG	ΔAGG	CDS; exon 58
***HTT*_SNP5**	**rs362306**	**A**	**G**	**3'UTR**
*HTT*_SNP6	rs362268	G	C	3'UTR
***HTT*_SNP7**	**rs362307**	**C**	**T**	**3'UTR**

Fig. 3 Assessing the specificity of HTT ddPCR assays. **a** HD cell lines used; D—differentiation. **b** Scheme of HTT transcript with marked CAG repeat tract (green triangle) and SNP variants identified in HD cell lines (red arrows—SNPs used in ddPCR assays; black arrows—other identified SNPs); light orange box—UTRs. Table presents all SNP variants identified in analyzed HD cell lines. Bolded SNPs were selected to be targets for ddPCR assays and they will be hereinafter referred to as HTT_SNP2, HTT_SNP5, and HTT_SNP7 in the text. CDS—coding sequence, Δ—deletion. RefSNP number according to the Single Nucleotide Polymorphism Database (dbSNP). **c** Results from ddPCR analysis of HTT_SNP2 assay specificity performed on seven samples with predefined ratios of WT/MUT plasmids (samples: I—100% WT and 0% MUT; II—90% WT and 10% MUT; III—75% WT and 25% MUT; IV—50% WT and 50% MUT; V—25% WT and 75% MUT; VI—10% WT and 90% MUT; VII—0% WT and 100% MUT). Precise values are indicated on WT/MUT bars ± poisson error. **d** Percentage ratio of HTT WT allele obtained with ddPCR HTT_SNP7 assay using gDNA from indicated HD cell lines. Three biological replicates were performed. These data are presented as means ± SD

*HTT*_SNP2 assay (Fig. 3c), as well as for *HTT*_SNP5 and *HTT*_SNP7 assays (Additional File 1: Fig. S1B and C). To determine the reliability of the HD ddPCR, we performed *HTT*_SNP7 assay (located in 3′UTR) on gDNA isolated from HD iPSCs, NSCs, and neurons. Amplification of *HTT* gDNA alleles from HD cells showed a canonical 50/50 WT/MUT ratio (Fig. 3d).

HTT WT/MUT ratio determination in HD cell lines

To verify allelic expression of the *HTT* gene, we examined HD iPSCs, NSCs, and neurons using *HTT*_SNP2, *HTT*_SNP5, and *HTT*_SNP7 ddPCR assays. Cellular identity of generated neurons was confirmed by immunostaining for expression of *DARPP32*, *TUJ1*, *MAP 2*, and *GAD67* neuronal markers (Additional File 2: Fig. S2). Allelic expression ratio in iPSCs averaged 68.7% WT/31.3% MUT (Fig. 4a; precise values for each *HTT* assay are given in Additional File 4: Fig. S4). In NSCs, we observed a slightly lower WT allele contribution, resulting in 64% WT/36% MUT ratio (Fig. 4b). In HD neurons, WT allele fraction dropped even further to 62.4% WT/37.6% MUT (Fig. 4c). The differences in allelic ratios between various HD cell lines were statistically significant (Fig. 4d). Comparing results obtained from HD and SCA3 cell lines, we documented greater differences between WT/MUT allele ratio in HD cells, than in SCA3 cells. We also noticed that the abundance of the MUT *HTT* allele increases with neural differentiation process (Fig. 4d).

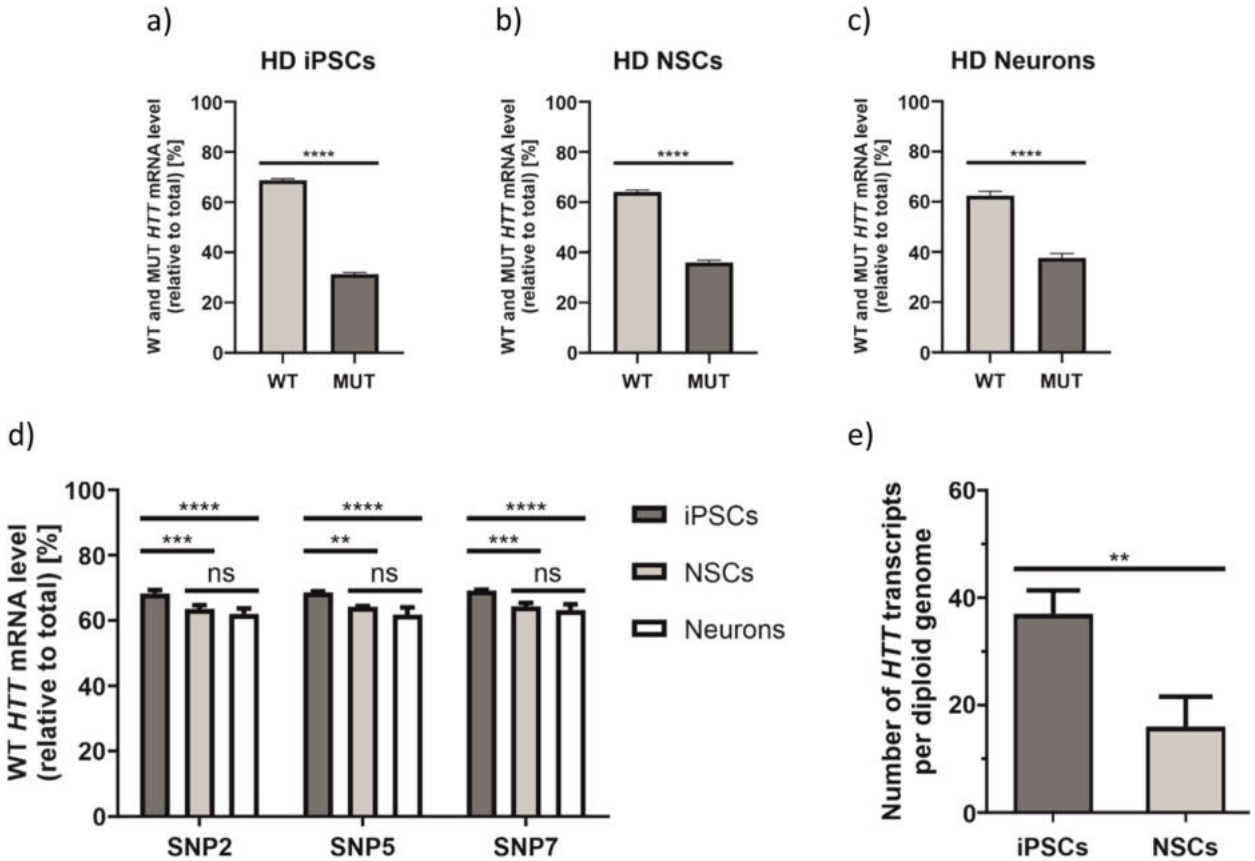


Fig. 4 Determination of endogenous WT and MUT HTT allele expression ratios in HD cell lines. **a–c** Results from ddPCR are presented as a mean WT and MUT HTT transcript allele abundance, calculated based on results from all three HTT assays (HTT_SNP2, HTT_SNP5 and HTT_SNP7), in iPSCs (***a***), NSCs (***b***), and neurons (***c***). These data were analyzed using unpaired *t* test. ***d*** Corresponding results to ***a–c*** presented for HTT_SNP2, HTT_SNP5, and HTT_SNP7 assays separately and analyzed with two-way ANOVA with Tukey's multiple comparison test for identification of any cell type-specific changes. In this graph, only WT allele abundance is presented, the MUT allele abundance equals a remainder to the sum of 100%). ***e*** Estimation of the total number (WT + MUT) of endogenous HTT transcripts per diploid genome using HTT_SNP5 assay. Data were analyzed using unpaired *t* test. For all experiments presented in this figure $n=3$. Two-tailed p value < 0.05 was considered significant and is depicted in the figure by: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. All data are presented as means ± SD. Individual data values are available in Additional File 13

To directly compare ddPCR results with information obtained from RNA-Seq, we extracted *HTT*-relevant reads from data obtained for HD iPSCs and NSC transcriptomes. We looked at *HTT* reads that span the sequences of three SNP variants selected for ddPCR design. The allelic expression ratio in iPSCs averaged 74.7% WT/25.3% MUT (Additional File 5: Fig. S5A), while in NSCs we obtained the ratio of 73.8% WT/26.2% MUT (Additional File 5: Fig. S5B), and these results were roughly consistent with ddPCR results. However, when each SNP region (SNP2, SNP5 or SNP7) was analyzed separately, we observed substantial differences in WT/MUT allele ratio, in both cell lines (Additional File 5: Fig. S5C-E), which generated higher error bars during statistical analysis. This implies that ddPCR is a more accurate and precise method for such approach.

Estimation of the number of endogenous HTT transcripts per diploid genome

To estimate an absolute number of *HTT* transcripts per diploid genome, we analyzed cDNA and gDNA from iPSCs and NSCs using *HTT*_SNP5 and *HTT*_SNP7 assays. As in SCA3 experiments, we divided the total number of *HTT* transcripts by the number of gDNA amplicons and, as a result, ~37 *HTT* transcripts in iPSC and ~16 *HTT* transcripts in an NSC were calculated for both *HTT*_SNP5 (Fig. 4e) and *HTT*_SNP7 assays (Additional File 6: Fig. S6). The difference between the number of *HTT* mRNAs in iPSCs and NSCs was statistically significant. Furthermore, the number of *HTT* transcripts calculated for NSCs corresponds well with our previous estimation obtained for this cell line using smFISH (~17 *HTT* mRNA molecules per cell) [28].

Exploration of WT/MUT HTT allelic expression after gene editing in iPSCs

We previously used a CRISPR-Cas9 method to correct the CAG expansion in HD iPSCs, and we generated isogenic control lines C39 and C105 [31]. Here, we inspected these lines to precisely determine expression of WT and corrected MUT (corrMUT) alleles. Additionally, we used a non-HD-related iPSC line derived from a healthy individual (C7522) that is also heterozygous for SNP5. When we examined gDNA of all three cell lines, we found nearly perfect 50/50 allelic ratio, which indicates that in the case of C39 and C105 lines CRISPR-Cas9 editing did not affect analyzed genomic sequences of the *HTT* gene (Fig. 5a). Our analyses of *HTT* allelic expression performed in C39 and C105 lines showed significantly altered expression in comparison with the initial iPSC line. While the initial iPSC line averaged 68.7% WT/31.3% MUT (Fig. 4a), C39 showed the average ratio of 75.9 WT/24.1% corrMUT (Fig. 5b; precise values for each *HTT* assay are given in Additional File 7: Fig. S7) and C105 revealed a ratio of 99.5 WT/0.5% corrMUT (Fig. 5c). All three SNP assay measurements for C39 and C105 were consistent (Additional File 7: Fig. S7), confirming almost total silencing of expression of the corrMUT allele in C105 line. Results obtained for C105 line suggest that there are off-target effects of gene editing that affected the expression of corrected allele, whereas the change in ratio of WT/corrMUT alleles in C39 line may result from CAG repeat tract shortening or, also, from some off-target effects. In control C7522 iPSC line, *HTT*_SNP5 assay showed 55.6% [WT-1]/44.4% [WT-2] ratio (Fig. 5d), suggesting more equal ratio in case of two WT alleles.

Silencing of endogenous HTT expression in HD NSCs

To verify the *HTT* ddPCR assay in assessing the efficacy of allele-selective gene silencing, we transfected HD NSCs with selected oligonucleotides, such as allele-selective miRNA-like oligonucleotide (A2) targeting expanded CAG tracts [32], non-allele-selective siRNA (siHTT) targeting *HTT* mRNA [33], and non-target control siRNA (siRL) targeting *Renilla* luciferase [28]. A2 oligo and siHTT reduced overall *HTT* expression by around 35% and around 50%, respectively, as measured by ddPCR (Fig. 6a). When we looked at *HTT* allelic expression, we observed equal downregulation of WT and MUT alleles in cells transfected with siHTT. However, silencing of the MUT allele expression was more prominent, as compared to WT allele, in cells transfected with A2 (Fig. 6b). This confirms that siHTT and A2 functions, in gene silencing mechanisms, in a non-allele-selective and an allele-selective manner, respectively [28], and highlights the utility of the HD ddPCR assay as a precise tool for determining the efficacy of transcript-reducing therapeutic strategies.

Examination of HTT transgenes in HD mice brain tissue

Determination of the HTT transgene copy number

To demonstrate the robustness of our HD ddPCR assay, we examined brain tissue from Hu128/21 mice, the HD mouse model created by crossing BAC21 and YAC128 mouse lines [34]. Hu128/21 mice carry *HTT* transgenes with heterozygous SNP variants for rs362331

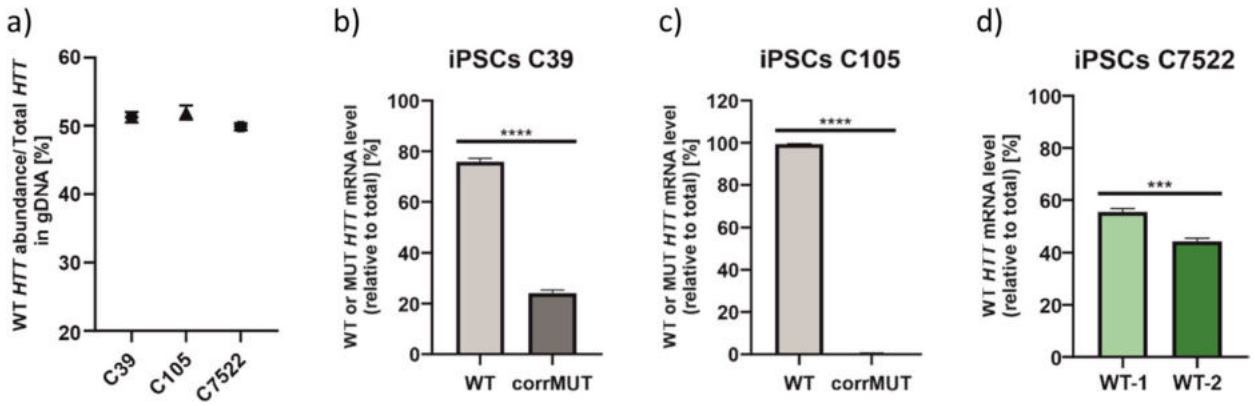


Fig. 5 Determination of endogenous HTT allele ratios in genetically corrected HD iPSCs and non-HD iPSCs. **a** Percentage ratio of HTT WT allele obtained with ddPCR HTT_SNP7 assay (or HTT_SNP5 regarding C7522 cells) using gDNA from indicated iPSC lines. ***b,c*** Results from ddPCR are presented as a mean WT or corrMUT HTT transcript allele abundance, calculated based on results from all three HTT assays (HTT_SNP2, HTT_SNP5, and HTT_SNP7), in genetically corrected C39 (***b***) and C105 (***c***). ***d*** Ratio of WT alleles of HTT transcript (WT-1 and WT-2) in control C7522 cell line. Data were analyzed using unpaired *t* test. For all experiments presented in this figure $n=3$. Two-tailed p value < 0.05 was considered significant and is depicted in the figure by: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. All data are presented as means ± SD. Individual data values are available in Additional File 13

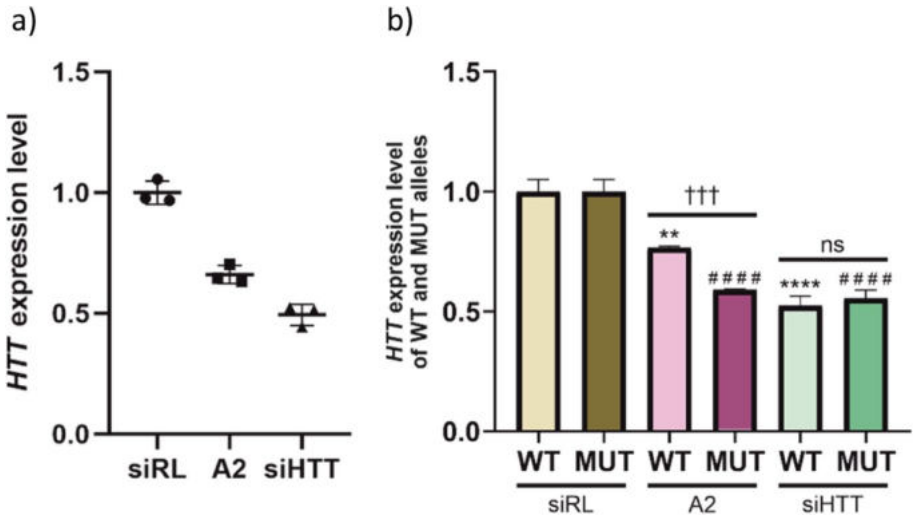


Fig. 6 Silencing of endogenous HTT expression in HD NSCs. **a** Endogenous HTT expression level after HD NSCs transfection with indicated oligonucleotides (siRL, A2, siHTT) assessed using ddPCR. ***b*** HTT WT and MUT allele abundance after transfection with indicated oligonucleotides. Results in ***a*** and ***b*** were normalized to HTT expression level in cells transfected with control siRNA targeting Renilla luciferase (siRL). Mean values from results obtained using HTT_SNP2, HTT_SNP5, and HTT_SNP7 are presented. Two-way ANOVA with Tukey's multiple comparison test was used to analyze changes in HTT expression level [WT (*) or MUT (#) allele], while unpaired *t* test was used to determine significance between WT and MUT HTT allele downregulation for each oligonucleotide (†). For all experiments presented in this figure $n=3$. Two-tailed p value < 0.05 was considered significant and is depicted in the figure by: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. All data are presented as means ± SD. Individual data values are available in Additional File 13

(*HTT*_SNP2) and rs362306 (*HTT*_SNP5), allowing us to determine copy number and expression of mutant and wild type alleles. In the first step, we analyzed gDNA isolated from Hu128/21 mice using *HTT*_SNP5 assay. As BAC and YAC *HTT* transgenes are present in various copy numbers, the assay in mice did not show 50/50 ratio. Instead, we documented a mean 64.5% WT/35.5% MUT ratio (Fig. 7a). Next, to determine precise transgene copy number in Hu128/21 mice, we performed a copy number variation (CNV) assay. We performed ddPCR for *Rpp30* as a reference and used *HTT*_SNP7 assay targeting *HTT* transgene. Since the rs362307 (*HTT*_SNP7) SNP variant is homozygous in alleles of the transgene in Hu128/21 mouse model, this assay recognizes total *HTT*, without WT/MUT discrimination. Our results showed that *HTT* transgene was around 8.5 times more abundant than a reference gene (represented by two copies); thus, each mouse had around 17 copies of *HTT* integrated into its genome (Fig. 7b). Taking WT/MUT ratio into consideration, Hu128/21 mice had around 11 copies of BAC21 transgene and around 6 copies of YAC128 transgene.

HTT WT/MUT ratio determination in selected brain tissues of two age groups of Hu 128/21 mice

Next, we analyzed WT/MUT allelic expression in striatum (STR), cortex (CTX), midbrain (MB), and olfactory bulb (OF) of presymptomatic and symptomatic Hu128/21 mice. In presymptomatic 4-month-old mice, we detected 63.4% WT/36.6% MUT ratio in STR, 64.5% WT/35.5% MUT ratio in CTX, 63.0% WT/37.0% MUT ratio in MB, and 66.5% WT/33.5% MUT ratio in OF (Fig. 7c; precise values for each *HTT* assay are given in Additional File 8: Fig. S8). When we quantified allelic expression in symptomatic 10-month-old mice, we obtained 65.9% WT/34.1% MUT ratio in STR, 65.8% WT/34.2% MUT ratio in CTX, 64.6% WT/35.4% MUT ratio in MB, and 69.1% WT/30.9% MUT ratio in OF (Fig. 7d). The allelic expression ratios matched the allelic distribution found using genomic DNA, suggesting that differences in transgene expression result directly from their copy numbers. Moreover, performed quantifications of *HTT* transcripts are consistent with reported WT/MUT huntingtin protein levels in the original paper describing Hu128/21 mice [34]. Using two age groups of Hu128/21 mice, we aimed to investigate any age-related changes in *HTT* transgene expression ratios in selected tissues (Fig. 7e, f). Two-way ANOVA performed on the data obtained for *HTT*_SNP2 assay and all four tissues considered/analyzed together showed that WT *HTT* mRNA abundance in 10-month-old mice was slightly higher, when compared to 4-month-old-mice (Fig. 7e), suggesting consistent increase of WT/MUT expression ratio in time in analyzed tissues. However, when all tissues were analyzed individually, post hoc test did not show a significance.

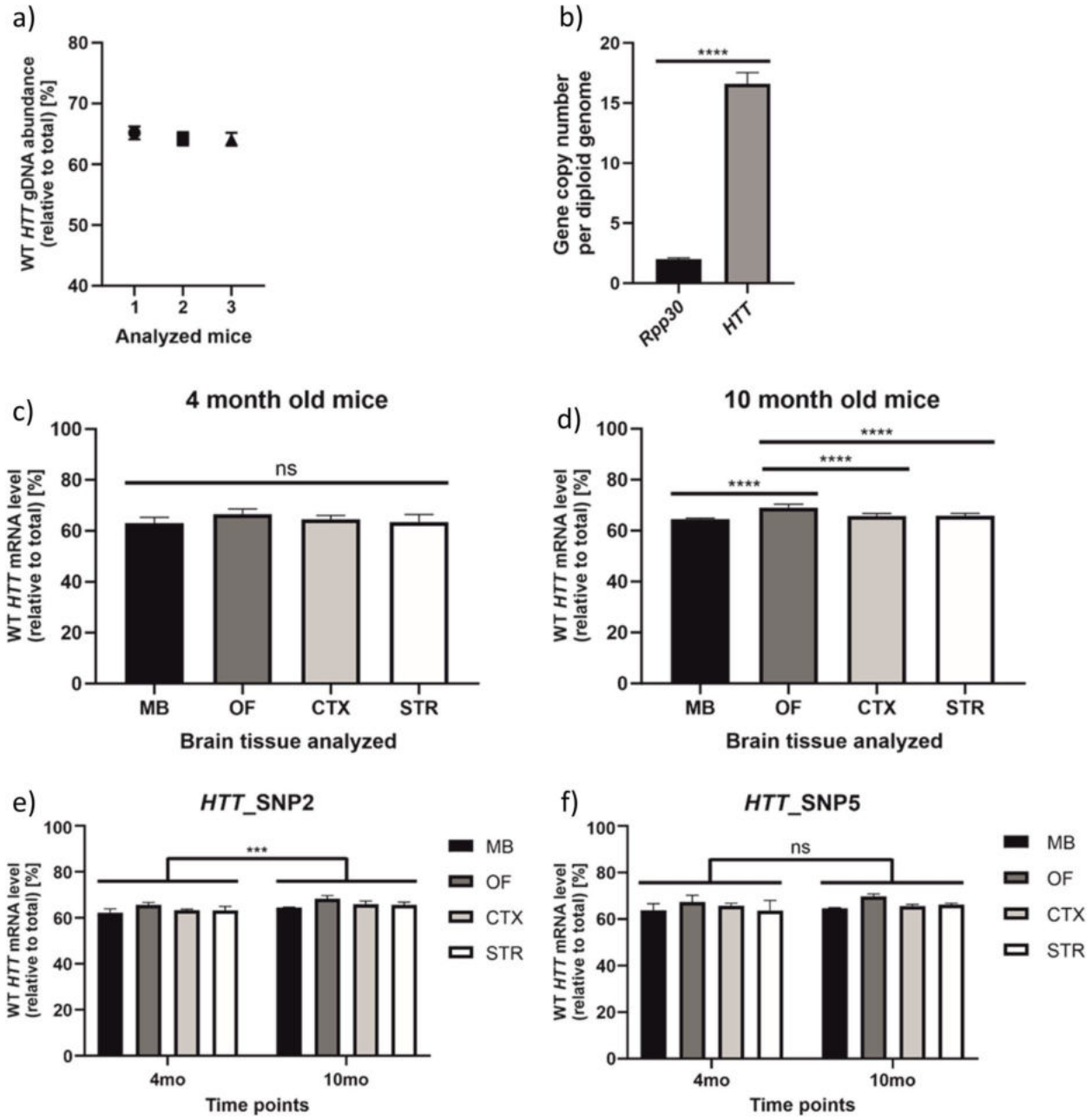


Fig. 7 Analyses on brain tissues of Hu128/21 mice aged 4 and 10 months. **a** Abundance of *HTT* transgene WT allele determined using *HTT*_SNP5 ddPCR assay on gDNA obtained from three Hu128/21 mice. **b** Copy number of *HTT* transgene in Hu128/21 mice, calculated using *HTT*_SNP7 assay, compared to the reference gene *Rpp30*. **c,d** *HTT* WT transcript abundance in midbrain (MB), olfactory bulb (OF), cortex (CTX), and striatum (STR) of 4-month-old (**c**) and 10-month-old (**d**) Hu128/21 mice. **e,f** Comparison of *HTT* WT transcript abundance in analyzed brain tissues between 4- and 10-month-old mice according to *HTT*_SNP2 assay (**e**), and according to *HTT*_SNP5 assay (**f**). Data were analyzed using unpaired t test (**b**), ordinary one-way ANOVA (**c, d**), or two-way ANOVA (**e, f**) with Sidak multiple comparison test. In **e**, we present results obtained by two-way ANOVA analysis, where all 4 tissues were grouped together and analyzed by the significance of the time point factor. For all experiments presented in this figure n=3. Two-tailed p value < 0.05 was considered significant and is depicted in the figure by: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. All data are presented as means ± SD. Individual data values are available in Additional File 13

There was no statistically significant time point-dependent changes of *HTT* allele expression ratio in test performed for particular tissue. However, according to the two-way ANOVA performed on the data obtained for *HTT*_SNP2 assay and all four tissues considered together, WT *HTT* mRNA abundance in 10-month-old

mice was slightly higher, as compared to 4-month-old mice (Fig. 7e), suggesting consistent increase of WT/MUT expression ratio in time in analyzed tissues.

Discussion

Three decades of extensive research have uncovered polyQ toxicity as a main driver of CAG repeat expansion diseases. However, a growing body of evidence suggests that CAG expanded transcripts also contribute to the pathogenesis and progression of these disorders [35, 36]. Investigation of RNA toxicity in dominantly inherited diseases is challenging since mRNA molecules transcribed from two nearly identical alleles of the causative genes cannot be easily distinguished. Since the polyQ disease genes are typically expressed at low levels, an ideal allele differentiating analytical method should demonstrate high accuracy, precision, sensitivity, and specificity. Because only a limited number of allele discrimination studies have been conducted in the context of transcripts related to polyQ diseases [37, 38], here, we developed an SNP-based ddPCR approach for allele-selective analyses of endogenous *HTT* and *ATXN3* transcripts. We performed a series of control amplifications using newly designed sets of primers and probes to show that our method is reliable and quantitative for analyzing low abundant transcripts with relatively complex GC-rich sequences. Analyses of the *ATXN3* and *HTT* transcripts in cells and tissues relevant to pathology of examined diseases, using independent SNP assays, demonstrated excellent reproducibility of transcript detection and discrimination.

The occurrence of differences in expression between two alleles of one gene is a known phenomenon [39], and several factors responsible for affecting mRNA allelic ratios were already identified using high-throughput analyses [40]. Interestingly, the presence of specific SNP variants itself can affect the expression level of alleles, e.g., through regulation of promoter activity, as shown for *HTT* [41], or changes in miRNA-binding sites in 3′UTR, which was recently shown for *ATXN3* expression [27]. Our allelic expression analysis in several HD and SCA3 cellular and mouse models revealed significant differences between WT and MUT alleles. These differences in selected patient-derived cell lines were generally lower in SCA3 cells (~54% WT/46% MUT) than in HD cells (62–69% WT/31–38% MUT) (Figs. 2e and 4d), but more cell lines should be used to draw general conclusions about *ATXN3* and *HTT* expression in patients' cells. Such broader analyses were performed on postmortem HD brain samples previously, using standard quantitative RT-PCR, rather than more precise ddPCR [37]. In that study, in the majority of tested samples, MUT *HTT* RNA was more abundant than the WT; however, a precise WT/MUT allele ratio seems to be an individual case. Another interesting issue to examine in the future is alternative transcript variants of both genes, *ATXN3* and *HTT*. In this research, we wanted to focus on the main variants of each transcript, but analyses on, e.g., *HTT*intron1 variant [14, 42] or *ATXN3* transcripts which does not contain exon 11 (which lead to the translation of ataxin3-a long and ataxin3-a short proteins) [26], might be interesting in terms of revealing new facts regarding the pathogenesis of both diseases.

For both cell lines, SCA3 and HD, we showed that the total number of the *ATXN3* and *HTT* transcripts, respectively, changes during neuronal differentiation in cell culture conditions (Figs. 2f and 4e). Estimations of examined transcripts per diploid genome showed diverse trends, which we confirmed using various calculation approaches. Moreover, these results corresponded well with smFISH-based detection of *ATXN3* and *HTT* in SCA3 fibroblasts and HD NSCs, respectively [28]. Interesting observations were made by Didiot et al., concerning normal *HTT* mRNA abundance and localization [43]. Not only they showed that the number of *HTT* mRNAs decreased with reprogramming of human fibroblasts into neuron-like cells, but also the localization of *HTT* mRNA changed—in neuron-like cells, *HTT* transcript was detected more often in the nuclear fraction, than in the cytoplasmic one. This is an interesting issue for further examination using both SCA3 and HD cells, together with discrimination of normal and mutant mRNA alleles. Additionally for *HTT*, we showed that abundance of MUT allele vs. WT allele increases with neuronal differentiation (Fig. 4d), what was also shown previously at a protein level in mouse cell lines obtained from HD YAC128 model [44, 45]. Again, these observations require confirmation in a larger set of cell lines, e.g., for investigation of processes that might be related to the phenomenon of neuronal vulnerability to mutations causing polyQ diseases [46]. However, these results support the idea that presented approach can be applied in studies aimed at the investigation of RNA biology of mutant transcripts leading to a better understanding of the pathogenesis of polyQ disorders [4, 5, 47–49]. Additionally, our approach could be also used for analyses of non-neuronal cells, as these cells were reported to be implicated in the pathogenesis of both SCA3 and HD [4, 50, 51]. Moreover, a joint characteristic for polyQ diseases is a regional vulnerability of specific brain regions to neurodegeneration. Hence, our assays could be used in determining whether specific cells (e.g., cortical neurons, Purkinje cells) show varied WT/MUT ratios or absolute numbers of transcripts per cell.

We applied our ddPCR-based approach in the evaluation of genetic engineering methods. The CRISPR-Cas9

technology provides a powerful tool for cell line generation, but off-target effects may affect gene expression and have an impact on the obtained results [52]. During the examination of potential off-target effects, an analysis performed on gDNA may provide only limited information about how the introduced modification affects the expression of the gene-of-interest. On the other hand, qPCR on cDNA is not as precise as ddPCR—since the amplification is exponential (meaning a 2-fold difference in number of molecules per cycle), qPCR is prone to even minor pipetting errors or incorrect choice of a reference gene [53]. Using previously generated isogenic control iPSC lines [31], we showed that quantitation using ddPCR can overcome this potential limitation. Additionally, based on RNA-Seq data from HD iPSCs and NSCs, we were able to estimate the ratio of WT/MUT alleles, but the accuracy and precision of ddPCR-based quantitation was much higher.

The analyses performed with the use of Hu128/21 mice tissues confirmed that the approach can be very useful for the investigation of polyQ disease-related transcripts in animal models. Unlike cells, transgenic Hu128/21 mice carried unequal distribution of WT and MUT gene copies already in gDNA. Interestingly, we showed differences in *HTT* mRNA levels originating from WT vs. MUT copies between analyzed tissues in the 10-month-old mice. Despite the fact that mostly striatal and cortical neurons are affected in HD [5], we have not observed significant differences in *HTT* mRNA WT/MUT ratios in those tissues, as compared to other two analyzed. Statistical analysis also showed that there was a slightly higher WT mRNA allele abundance in 10-month-old mice in comparison to 4-month-old animals. This observation is consisted with results from human HD brain tissues, where lower abundance of WT mRNA allele (relative to MUT allele) was reported in samples from patients at early disease stages, as compared to samples from patients at late disease stages [37]. However, more testing is required to verify whether those observations are general or model-specific. So far, time course of mutant and total HTT protein levels was investigated in detail in BACHD rats, showing that the ratio between mutant and total huntingtin can significantly change in time [54].

There is still no available treatment for polyQ diseases. The experimental approaches which are considered as the most promising (as the most efficient and safe) for HD and SCA3 are based on allele-selective targeting of mutant alleles [55–58]. Preferential targeting of mutant transcripts requires precise evaluation of efficiency and selectivity of the used strategy, e.g., through quantitative analysis of WT and MUT protein levels [59–61]. However, protein assays usually require more tissue or CSF, which is often limited. The solution can be a highly sensitive transcript assay, like the one described in this study. Since measurements of the mutant gene expression level are also considered as important biomarker in polyQ diseases [37, 62, 63], our approach can be considered as a powerful tool in clinical research.

Conclusions

We adapted a SNP-based ddPCR method for the evaluation of WT/MUT allele expression ratio in SCA3 and HD cells, as well as in the HD mouse model. Our results clearly show that the allele-specific ddPCR methods presented in this work is suitable for the detection of slight variations in the expression levels of tested alleles in both cell cultures and mice tissues implicated in polyQ diseases. We examined various applications where our approach can be used and, most importantly, we showed its usefulness for accurate and precise analysis of allelic expression during cell differentiation, and after gene editing, as well as in assessing the efficacy of polyQ therapies.

Methods

Cell lines and cell culture

SCA3 fibroblasts (GM06153, 16/67 CAG repeats in *ATXN3*) were obtained from the NIGMS Repository (Coriell Institute for Medical Research). They were cultivated in Minimum Essential Medium (Gibco), containing 10% fetal bovine serum (Biowest), a penicillin-streptomycin solution (Sigma-Aldrich), and 2mM L-glutamine (Sigma-Aldrich). SCA3 iPSCs were generated (from GM06153 fibroblasts) previously [64]. HD iPSCs (ND42222, 19/109 CAG repeats in *HTT*) were obtained from the NIGMS Repository. Additional cell lines were used: iPSC clonal lines (C39 and C105) previously generated via CRISPR-Cas9-based modification of the ND2222 line [31], and a non-HD-related iPSC line (C7522) generated from a healthy individual-derived fibroblast line (GM07522) obtained from the NIGMS Repository. iPSCs were cultured in StemFlex (Gibco) medium on Geltrex (Gibco). NSCs were grown in STEMdiff Neural Progenitors Medium (NPM) (STEMCELL Technologies).

Differentiation of HD and SCA3 iPSCs to NSCs

iPSCs were differentiated to NSCs using the STEMdiff SMADi Neural Induction Kit (STEMCELL Technologies) and a monolayer protocol, following the manufacturer's instructions. Briefly, iPSCs were grown on a 6-well plate until 70–80% confluence was reached, and then dissociated to single cells by incubation with Accutase (STEMCELL Technologies) for 4 min. Cells were counted using TC20 Automated Cell Counter (Bio-Rad) and resuspended at 1×10^6 cells/ml density for seeding in STEMdiff Neural Induction Medium with SMADi and 10 nM

Y-27632 (all from STEMCELL Technologies). After third passage, they were grown in STEMdiff NPM.

Differentiation of SCA3 NSCs to neurons

SCA3 NSCs were differentiated to neurons according to protocol published by Hansen et al. [65] with slight modifications. Briefly, NSC were grown until full confluence and then dissociated to single cells by incubation with Accutase. Collected cells were passaged in 1:10 ratio onto Geltrex-coated 6-well and 12-well (for ICC) plates with NPM and 10 nM Y-27632. Next-day NPM was exchanged for neural maturation medium N2B27 with 20 ng/mL recombinant human BDNF and 10ng/mL recombinant human GDNF (both STEMCELL Technologies) and 300 ng/μL cAMP (Sigma-Aldrich). Medium was changed every 2 days for 3 weeks.

Differentiation of HD NSCs to MSNs

HD NSC neural differentiation was performed according to protocol published by M. Fjodorova and M. Li [66] with slight modifications. Briefly, NSCs were grown in STEMdiff NPM until full confluence and then dissociated to single cells by incubation with Accutase. Collected cells were passaged in 1:5 ratio onto a Geltrex-coated 6-well plate with NPM and 10 nM Y-27632. Next-day NPM was exchanged for LGE pattering medium with 25 ng/mL recombinant human Activin A (STEMCELL Technologies). Medium was changed every day until day 10 when LGE progenitors were passaged for terminal differentiation into medium spiny neurons (MSNs): cells were passaged with Accutase in ratio 1:5 onto Geltrex precoated 6-well and 12-well plates (for ICC) in LGE pattering medium and 10 nM Y-27632. Next-day LGE pattering medium was exchanged for a terminal differentiation medium N2B27 with 10 ng/mL recombinant human BDNF and 10 ng/mL recombinant human GDNF (both STEMCELL Technologies). Medium was changed every 2 days for 3 weeks.

Immunocytochemistry (ICC)

NSC and neuronal cells were fixed firstly for 5 min by adding 2% PFA directly to a cell culture medium and then gently washed in PBS followed by fixation in 4% PFA in PBS for 10 min. Cells were permeabilized with 0.5% Tween, blocked in 1% bovine serum albumin, and incubated with primary antibodies and fluorescent-dye conjugated secondary antibodies (listed in Additional File 9: Table S1). DAPI was used for DNA staining. Images were captured with a Leica DMI6000 microscope (Additional File 10: Fig. S2).

HD mice

Hu128/21 HD mice model was described previously [34]. Briefly, the Hu128/21 mouse model was created by crossing Hdh−/− BAC21 mice (Hu21) and Hdh−/− YAC128 mice (Hu128), and then genotyped. Mice were housed in the animal facility of the Center for Advanced Technology (CAT) in Poznań and kept under standard conditions with a 12-h light/dark cycle with water and food ad libitum. All procedures and handling of the animals were performed to minimize stress to the animals and according to a protocol approved and monitored by the Local Ethical Commission for Animal Experiments in Poznan (Poland). Mice were sacrificed and brain tissues harvested and snap frozen in liquid nitrogen and stored at −80°C. Number and age of mice used in all experiments are reported in the accompanying figure legends.

SNP variant identification

For SNP variant identification, the TOPO XL-2 PCR Cloning Kit (Invitrogen) was used. Briefly, RNA isolated from SCA3 and HD cell lines (iPSCs) was reversely transcribed using High-Capacity cDNA Reverse Transcription kit (Applied Biosystems) with random hexamer primers (for *ATXN3*) or SuperScript IV Reverse Transcriptase (Thermo Fisher) with a gene-specific primer (HTT_RT) (for *HTT*). Then, cDNA served as a template in PCR reactions using Platinum SuperFi polymerase included in a cloning kit. In case of *HTT*, PCRs were performed using PrimeSTAR GXL DNA Polymerase (Takara Bio). Primers in those PCR reactions were designed in a way to amplify whole CDSs, with fragments of both 5′ and 3′UTRs (ATXN3_Fwd and ATXN3_Rev; HTT_Fwd and HTT_Rev, sequences given in Additional File 10: Table S2). PCR products were then cloned into pCR-XL-2-TOPO vector. Generated vectors were transformed into bacterial cells provided with the kit. Transformants were analyzed through colony PCR and restriction enzyme digestion to select clones with either WT or MUT versions of analyzed genes. Positive clones were then sequenced, and SNPs allowing for discrimination between WT and MUT alleles, after additional confirmation on genomic DNA (gDNA), were identified. All identified SNPs are listed in Figs. 1b and 3b.

RNA isolation and reverse transcription

For RNA isolation, the Total RNA Zol-Out D kit (A&A Biotechnology) was used, according to the manufacturer's protocol. In case of cells, cell pellets were suspended in 800 μL of TRI Reagent (Thermo Fisher), while mouse tissues were homogenized in 300 μL of TRI Reagent. The concentration of isolated total RNA was calculated by measurement at 260 nm using DeNovix

spectrophotometer. One microgram of RNA was then used for reverse transcription using the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems) with random hexamer primers.

Genomic DNA isolation

gDNA was isolated using the Genomic DNA Isolation kit (Norgen Biotek Corp.) according to the manufacturer's protocols designed for cells growing in a monolayer and animal tissues. The concentration of isolated gDNA was calculated by measurement at 260 nm using the DeNovix spectrophotometer.

Droplet digital PCR

Among various ddPCR applications, Rare Mutation and Sequence Detection were chosen. To specifically target sequences differing by only one nucleotide, two probes labelled with different fluorophores were designed together with a pair of primers. Because of the high dilution of the sample, probes bind to templates carrying specific SNP variants with a high specificity. After the readout based on the fluorescence, data were analyzed using Poisson statistics to determine the concentration of template carrying each SNP variant (and its flanking sequence) in the original sample. ddPCRs were prepared using DG8 cartridges and gaskets, Droplet Generation Oil for Probes, and ddPCR Supermix for Probes (no dUTP) (Bio-Rad), and they were performed on the QX200 Droplet Digital PCR System (Bio-Rad), according to the manufacturer's protocols and established MIQE guidelines for ddPCR [67]. ddPCR assays targeting identified SNP variants were designed by a Bio-Rad's technician and their sequences are unknown; however, their assay IDs are listed in Additional File 10: Table S2. As a reference, ddPCR assay targeting β-actin was used, while for determination of *HTT* transgene copy number, an assay targeting mouse *Rpp30* was used.

NSC transfection with oligonucleotides

All siRNA oligonucleotides (Additional File 11: Table S3) were synthesized by FutureSynthesis, dissolved in water to 100 µM concentration, and stored at −80 °C. To obtain 20 µM duplexes, sense and antisense strands were diluted in water, heated for 1 min in 90 °C, and kept for gradual cooling at room temperature for 45 min. HD NSCs were then transfected using 2 µL of siPORT Amine (Ambion) and 100 nM oligonucleotides per well of a 6-well plate in 1 mL of complete medium. After 3 h, the medium was replaced with a fresh one, and after the next 24 h, cells were harvested and total RNA was isolated and then used for reverse transcription and ddPCRs, as previously described.

RNA-Seq

RNA-Seq was performed in the CeNT's Genomics Core Facility in the Centre of New Technologies, University of Warsaw, Poland. Total RNA isolated from HD iPSCs and NSCs after ribodepletion was sequenced using NovaSeq 6000 system with a pair-end 2x100 cycles procedure. Fastq files were mapped to the reference genome (GENCODE release 42, GRCh38.p13) using STAR (v. 2.7.10b) algorithm. Only the reads mapped to the forward strand were subsetted (samtools 1.16.1). The statistics for reads mapped to the investigated SNPs in *HTT* were retrieved using samtools mpileup (QC parameters: -q 20 -Q 20).

Statistical analysis

Three biological replicates were included for each experiment. Analyses of obtained results were performed using GraphPad Prism software (version 8.0.1.). Depending on the experimental setup, specific statistical tests were used and are indicated in figure legends. All data are presented as means. Error bars always represent standard deviations.

Abbreviations

PolyQ	Polyglutamine
ddPCR	Droplet digital PCR
WT	Wild type
MUT	Mutant
HTT	Huntingtin
ATXN3	Ataxin-3
HD	Huntington's disease
SCA3	Spinocerebellar ataxia type 3

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-023-01515-3.

Additional file 1: Fig. S1. Analyses of *ATXN3* and *HTT* assays specificity. Results from ddPCR assays performed on seven samples with predefined ratios of WT/MUT *ATXN3* (A) and *HTT* (B-C) plasmids (I—100% WT and 0% MUT; II—90% WT and 10% MUT; III—75% WT and 25% MUT; IV—50% WT and 50% MUT; V—25% WT and 75% MUT; VI—10% WT and 90% MUT; VII—0% WT and 100% MUT) using ATXN3_SNP5 (A), HTT_SNP5 (B) and HTT_SNP7 assays (C). Mean/Precise values are indicated on WT/MUT bars ± poisson error.

Additional file 2: Fig. S2. Immunocytochemistry performed on SCA3 and HD neurons. Exemplary images of immunofluorescent staining of SCA3 neurons for MAP 2 and TUJ1, and HD neurons for DARPP32, TUJ1, MAP 2 and GAD67. DAPI was used for nuclei staining.

Additional file 3: Fig. S3. Precise values describing WT/MUT ratio of *ATXN3* in selected SCA3 cells. A) Table with mean %WT and %MUT values obtained using four types of SCA3 cell lines and two SNP assays (mean values from three biological replicates). B-E) Results from ddPCR presented as a mean WT/MUT *ATXN3* transcript allele abundance obtained using ATXN3_SNP2 and ATXN3_SNP5 assays in fibroblasts (B), iPSCs (C), NSCs (D) and neurons (E). To determine statistical significance of differences between WT and MUT allele abundance data were analyzed using unpaired t test. For all experiments presented in this figure $n=3$. Two-tailed p value < 0.05 was considered significant and is depicted in the

figure by: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. All data are presented as means ± SD.

Additional file 4: Fig. S4. Precise values describing WT/MUT ratio of *HTT* in selected HD cells. A) Table with mean %WT and %MUT values obtained using three types of HD cell lines and three SNP assays (mean values from three biological replicates). B-D) Results from ddPCR presented as a mean WT and MUT *HTT* transcript allele abundance obtained using HTT_SNP2, HTT_SNP5 and HTT_SNP7 assays in iPSCs (B), NSCs (C) and neurons (D). To determine statistical significance of differences between WT and MUT allele abundance, data were analyzed using unpaired t test. For all experiments presented in this figure $n=3$. Two-tailed p value < 0.05 was considered significant and is depicted in the figure by: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. All data are presented as means ± SD.

Additional file 5: Fig. S5. RNA-Seq analysis of *HTT* reads performed on HD iPSCs and NSCs. A-B) Results from RNA-Seq are presented as a mean WT and MUT *HTT* transcript allele abundance, calculated based on reads from three *HTT* SNP regions (SNP2, SNP5 and SNP7), in iPSCs (A) and NSCs (B). These data were analyzed using unpaired t test. C) Table summarizing precise RNA-Seq results. D-E) Precise results from RNA-Seq presented as a WT and MUT *HTT* transcript allele abundance based on reads from three, particular SNP regions (SNP2, SNP5 and SNP7) in iPSCs (D) and NSCs (E). Data were analyzed using unpaired t test. For all experiments presented in this figure $n=3$. All data are presented as means ± SD. Individual data values are available in Additional File 13.

Additional file 6: Fig. S6. Estimation of the total number of endogenous *HTT* transcripts per diploid genome. Results were obtained using *HTT*_SNP7 assay and cDNA from HD iPSCs and NSCs. Data were analyzed using unpaired t test. For all experiments presented in this figure $n=3$. Two-tailed p value < 0.05 was considered significant and is depicted in the figure by: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. All data are presented as means ± SD.

Additional file 7: Fig. S7. Precise values describing WT/MUT ratio of *HTT* in isogenic controls to HD cells. A) Table with mean %WT and %corrMUT values obtained for lines and presented SNP assays (mean values from three biological replicates). B-C) Results from ddPCR presented as a mean WT/MUT *HTT* transcript allele abundance obtained using HTT_SNP2, HTT_SNP5 and HTT_SNP7 assays in C39 (B) and C105 (C). To determine statistical significance of differences between WT and MUT allele abundance, data were analyzed using unpaired t test. For all experiments presented in this figure $n=3$. Two-tailed p value < 0.05 was considered significant and is depicted in the figure by: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. All data are presented as means ± SD. Error bars in the figure represent standard deviations.

Additional file 8: Fig. S8. Precise values describing WT/MUT ratio of *HTT* transgene in Hu128/21 mice. A) Table with mean %WT and %MUT values obtained using mice brain tissues and two ddPCR assays (mean values from three biological replicates). B-C) Results from ddPCR presented as a mean WT/MUT *HTT* transcript allele abundance obtained using HTT_SNP2 and HTT_SNP5 assays in 4-month-old (B) and 10-month-old mice (C). Data were analyzed using two-way ANOVA with Tukey's multiple comparison test. For all experiments presented in this figure $n=3$. Two-tailed p value < 0.05 was considered significant and is depicted in the figure by: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. All data are presented as means ± SD. Error bars in the figure represent standard deviations.

Additional file 9: Table S1. Table listing all primers and ddPCR assays used in this research with their description.

Additional file 10: Table S2. Sequences of oligonucleotides used for silencing of endogenous *HTT* in HD NSCs.

Additional file 11: Table S3. A list of primary and secondary antibodies with their dilutions used in immunocytochemistry.

Additional file 12 Selected reads from RNA-Seq analysis. Reads covering *HTT* SNPs of interest in each sample (3 samples from HD iPSCs—iH222HD and 3 samples from HD NSCs—nH222HD) selected from aligned bam files (samtools view) and converted to the fastq format (samtools fastq).

Additional file 13. Individual data values for selected figures.

Acknowledgements
All ddPCR analyses were performed in the Laboratory of Single Cell Analyses, IBCH PAS. Microscopic images were obtained in the Laboratory of Subcellular Structures Analyses, IBCH PAS. The authors would like to acknowledge Włodzimierz J. Krzyżosiak as an initiator of a general concept of the study, and thank Marta Pietras, Maria Ciesiołka, and Bartosz Nowak for contributions to the preliminary analyses of this study, and Edyta Kościańska for comments on the manuscript.

Authors' contributions
Conceptualization: AF, AC1, and PMS. Identification of SNP variants: PJ, AC1, and GF. ddPCR experiments: PJ. Cell culture: EK, AC2. Immunocytochemistry: EK. Analysis of *HTT* RNA-Seq reads: GA. NSC transfection: PMS, EK, and PJ. Mice handling and dissections: MS, ZK-P, and MF. Mouse model sharing and advisory: NSC, MRH. Manuscript writing: PJ, AF, and PMS, with the input and revision from all authors. All authors read and approved the final manuscript. AC1—Adam Ciesiołka; AC2—Agata Ciołak.

Funding
This work was supported by Grants from the National Science Centre, Poland [2015/19/B/NZ2/02453 - cell culture experiments, 2021/41/B/NZ3/03803 - ddPCR assays from mouse tissue] and the consortium Grant from the National Centre for Research and Development in frame of European Research Projects On Rare Diseases (JTC 2017) [ERA-NET-E-RARE-3/III/TreatPolyQ/08/2018] - mice breeding and tissue dissection].

Availability of data and materials
All data generated or analyzed during this study are included in this published article and its supplementary information files. Analyzed RNA-Seq reads are provided as Additional file 12. Individual data values from ddPCR experiments are provided as Additional file 13.

Declarations

Ethics approval and consent to participate
All procedures performed on animals were approved and monitored by the Local Ethical Commission for Animal Experiments in Poznan (Poland) (approval number 64/2018).

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Medical Biotechnology, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland. [2]Department of Molecular Neurobiology, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland. [3]Centre for Molecular Medicine and Therapeutics, BC Children's Hospital Research Institute, Department of Medical Genetics, University of British Columbia, Vancouver, BC V5Z 4H4, Canada.

Received: 6 September 2022 Accepted: 17 January 2023
Published online: 01 February 2023

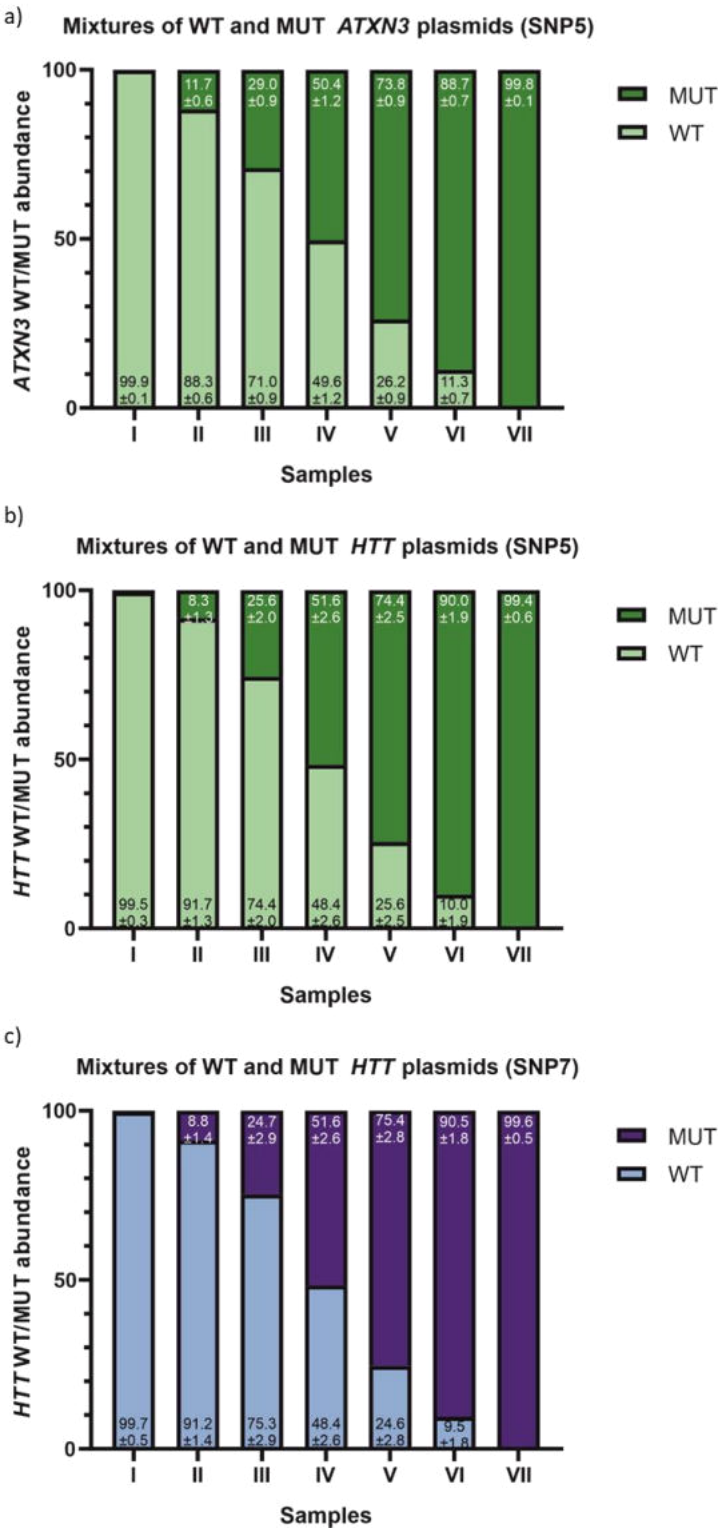
References

1. Lieberman AP, Shakkottai VG, Albin RL. Polyglutamine repeats in neurodegenerative diseases. Annu Rev Pathol Mech Dis. 2019;14:1–27.
2. Bunting EL, Hamilton J, Tabrizi SJ. Polyglutamine diseases. Curr Opin Neurobiol. 2022;72:39–47.
3. Stoyas CA, La Spada AR. The CAG–polyglutamine repeat diseases: a clinical, molecular, genetic, and pathophysiologic nosology. In: Handbook of Clinical Neurology: Elsevier; 2018. p. 143–70.
4. McLoughlin HS, Moore LR, Paulson HL. Pathogenesis of SCA3 and implications for other polyglutamine diseases. Neurobiol Dis. 2020;134:104635.

5. Tabrizi SJ, Flower MD, Ross CA, Wild EJ. Huntington disease: new insights into molecular pathogenesis and therapeutic opportunities. Nat Rev Neurol. 2020;16:529–46.
6. Yushchenko T, Deuerling E, Hauser K. Insights into the aggregation mechanism of PolyQ proteins with different glutamine repeat lengths. Biophys J. 2018;114:1847–57.
7. Saunders HM, Bottomley SP. Multi-domain misfolding: understanding the aggregation pathway of polyglutamine proteins. Protein Eng Design Select. 2009;22:447–51.
8. Da Silva JD, Teixeira-Castro A, Maciel P. From pathogenesis to novel therapeutics for spinocerebellar ataxia type 3: evading potholes on the way to translation. Neurotherapeutics. 2019;16:1009–31.
9. Matos CA, de Almeida LP, Nóbrega C. Machado–Joseph disease/spinocerebellar ataxia type 3: lessons from disease pathogenesis and clues into therapy. J Neurochem. 2019;148:8–28.
10. Nalavade R, Griesche N, Ryan DP, Hildebrand S, Krauß S. Mechanisms of RNA-induced toxicity in CAG repeat disorders. Cell Death Dis. 2013;4:752.
11. Chan HYE. RNA-mediated pathogenic mechanisms in polyglutamine diseases and amyotrophic lateral sclerosis. Front Cellular Neurosci. 2014;8:431.
12. Martí E. RNA toxicity induced by expanded CAG repeats in Huntington's disease. Brain Pathol. 2016;26:779–86.
13. Fiszer A, Krzyzosiak WJ. RNA toxicity in polyglutamine disorders: concepts, models, and progress of research. J Mol Med. 2013;91:683–91.
14. Fienko S, Landles C, Sathasivam K, McAteer SJ, Milton RE, Osborne GF, et al. Alternative processing of human HTT mRNA with implications for Huntington's disease therapeutics. Brain. 2022;145:4409–24.
15. Jazurek M, Ciesiolka A, Starega-Roslan J, Bilinska K, Krzyzosiak WJ. Identifying proteins that bind to specific RNAs - focus on simple repeat expansion diseases. Nucleic Acids Res. 2016;44:9050–70.
16. An Y, Chen ZS, Yin H, Chan E, Chi J, Ngo K. Molecular insights into the interaction of CAG trinucleotide RNA repeats with nucleolin and its implication in polyglutamine diseases. 2022;50:7655–68.
17. Ly S, Didiot M-C, Ferguson CM, Coles AH, Miller R, Chase K, et al. Mutant huntingtin messenger RNA forms neuronal nuclear clusters in rodent and human brains. Brain Commun. 2022;4:fcac248.
18. Jain A, Vale RD. RNA phase transitions in repeat expansion disorders. Nature. 2017;546:243–7.
19. Joachimiak P, Ciesiołka A, Figura G, Fiszer A. Implications of Poly(A) tail processing in repeat expansion diseases. Cells. 2022;11:677.
20. Fiszer A, Krzyzosiak WJ. Oligonucleotide-based strategies to combat polyglutamine diseases. Nucleic Acids Res. 2014;42:6787–810.
21. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature. 2012;489:391–9.
22. Anderson BR, Jensen ML, Hagedorn PH, Little SC, Olson RE, Ammar R, et al. Allele-selective knockdown of MYH7 using antisense oligonucleotides. Mol Ther - Nucleic Acids. 2020;19:1290–8.
23. Dodd DW, Gagnon KT, Corey DR. Digital quantitation of potential therapeutic target RNAs. Nucleic Acid Ther. 2013;23:188–94.
24. Kourkouta E, Weij R, González-Barriga A, Mulder M, Verheul R, Bosgra S, et al. Suppression of mutant protein expression in SCA3 and SCA1 mice using a CAG repeat-targeting antisense oligonucleotide. Mol Ther - Nucleic Acids. 2019;17:601–14.
25. Wojciechowska M, Sobczak K, Kozlowski P, Sedehizadeh S, Wojtkowiak-Szlachcic A, Czubak K, et al. Quantitative methods to monitor RNA biomarkers in myotonic dystrophy. Sci Rep. 2018;8:5885.
26. Weishäupl D, Schneider J, Pinheiro BP, Ruess C, Dold SM, von Zweydorf F, et al. Physiological and pathophysiological characteristics of ataxin-3 isoforms. J Biol Chem. 2019;294:644–61.
27. Melo ARV, Raposo M, Ventura M, Martins S, Pavão S, Alonso I, et al. Genetic Variation in ATXN3 (Ataxin-3) 3′UTR: insights into the downstream regulatory elements of the causative gene of Machado-Joseph disease/spinocerebellar ataxia type 3. Cerebellum. 2022. https://doi.org/10.1007/s12311-021-01358-0.
28. Ciesiolka A, Stroynowska-Czerwinska A, Joachimiak P, Ciolak A, Kozlowska E, Michalak M, et al. Artificial miRNAs targeting CAG repeat expansion in ORFs cause rapid deadenylation and translation inhibition of mutant transcripts. Cell Mol Life Sci. 2021;78:1577–96.
29. Kay C, Collins JA, Caron NS, de Agostinho LA, Findlay-Black H, Casal L, et al. A comprehensive haplotype-targeting strategy for allele-specific HTT suppression in Huntington disease. Am J Hum Genet. 2019;105:1112–25.
30. Pfister EL, Kennington L, Straubhaar J, Wagh S, Liu W, DiFiglia M, et al. Five siRNAs targeting three SNPs may provide therapy for three-quarters of Huntington's disease patients. Curr Biol. 2009;19:774–8.
31. Dabrowska M, Ciolak A, Kozlowska E, Fiszer A, Olejniczak M. Generation of new isogenic models of Huntington's disease using CRISPR-Cas9 technology. Int J Mol Sci. 2020;21:1854.
32. Fiszer A, Olejniczak M, Galka-Marciniak P, Mykowska A, Krzyzosiak WJ. Self-duplexing CUG repeats selectively inhibit mutant huntingtin expression. Nucleic Acids Res. 2013;41:10426–37.
33. Wang YL, 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;53:241–9.
34. Southwell AL, Skotte NH, Villanueva EB, Østergaard ME, Gu X, Kordasiewicz HB, et al. A novel humanized mouse model of Huntington disease for preclinical development of therapeutics targeting mutant huntingtin alleles. Hum Mol Genet. 2017;26:1115–32.
35. Heinz A, Nabariya DK, Krauss S. Huntingtin and its role in mechanisms of RNA-mediated toxicity. Toxins. 2021;13:487.
36. Krzyzosiak WJ, Sobczak K, Wojciechowska M, Fiszer A, Mykowska A, Kozlowski P. Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. Nucleic Acids Res. 2012;40:11–26.
37. Liu W, Chaurette J, Pfister EL, Kennington LA, Chase KO, Bullock J, et al. Increased steady-state mutant huntingtin mRNA in Huntington's disease brain. J Huntingtons Dis. 2013;2:491–500.
38. Shin A, Shin B, Shin JW, Kim KH, Atwal RS, Hope JM, et al. Novel allele-specific quantification methods reveal no effects of adult onset CAG repeats on HTT mRNA and protein levels. Hum Mol Genet. 2017;26:1258–67.
39. Buckland PR. Allele-specific gene expression differences in humans. Hum Mol Genet. 2004;13:R255–60.
40. Smith RM, Webb A, Papp AC, Newman LC, Handelman SK, Suhy A, et al. Whole transcriptome RNA-Seq allelic expression in human brain. BMC Genom. 2013;14:1–15.
41. Becanovic K, Nørremølle A, Neal SJ, Kay C, Collins JA, Arenillas D, et al. A SNP in the HTT promoter alters NF-κB binding and is a bidirectional genetic modifier of Huntington disease. Nat Neurosci. 2015;18:807–16.
42. Sathasivam K, Neueder A, Gipson TA, Landles C, Benjamin AC, Bondulich MK, et al. Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. Proc Natl Acad Sci U S A. 2013;110:2366–70.
43. Didiot MC, Ferguson CM, Ly S, Coles AH, Smith AO, Bicknell AA, et al. Nuclear localization of huntingtin mRNA is specific to cells of neuronal origin. Cell Rep. 2018;24:2553–2560.e5.
44. Szlachcic WJ, Wiatr K, Trzeciak M, Figlerowicz M, Figiel M. The generation of mouse and human huntington disease iPS cells suitable for In vitro studies on huntingtin function. Front Mol Neurosci. 2017;10:253.
45. Szlachcic WJ, Switonski PM, Krzyzosiak WJ, Figlerowicz M, Figiel M. Huntington disease iPSCs show early molecular changes in intracellular signaling, the expression of oxidative stress proteins and the p53 pathway. DMM Dis Model Mech. 2015;8:1047–57.
46. Saxena S, Caroni P. Selective neuronal vulnerability in neurodegenerative diseases: from stressor thresholds to degeneration. Neuron. 2011;71:35–48.
47. Jimenez-Sanchez M, Licitra F, Underwood BR, Rubinsztein DC. Huntington's disease: mechanisms of pathogenesis and therapeutic strategies. Cold Spring Harb Perspect Med. 2017;7:1–22.
48. Jazurek-Ciesiolka M, Ciesiolka A, Komur AA, Urbanek-Trzeciak MO, Krzyzosiak WJ, Fiszer A. RAN translation of the expanded CAG repeats in the SCA3 disease context. J Mol Biol. 2020;432:166699.
49. Bañez-Coronel M, Ayhan F, Tarabochia AD, Zu T, Perez BA, Tusi SK, et al. RAN translation in Huntington disease. Neuron. 2015;88:667–77.
50. Onur TS, Laitman A, Zhao H, Keyho R, Kim H, Wang J, et al. Downregulation of glial genes involved in synaptic function mitigates Huntington's disease pathogenesis. Elife. 2021;10:1–30.
51. Wilton DK, Stevens B. The contribution of glial cells to Huntington's disease pathogenesis. Neurobiol Dis. 2020;143:104963.
52. Sledzinski P, Dabrowska M, Nowaczyk M, Olejniczak M. Paving the way towards precise and safe CRISPR genome editing. Biotechnol Adv. 2021;49:107737.

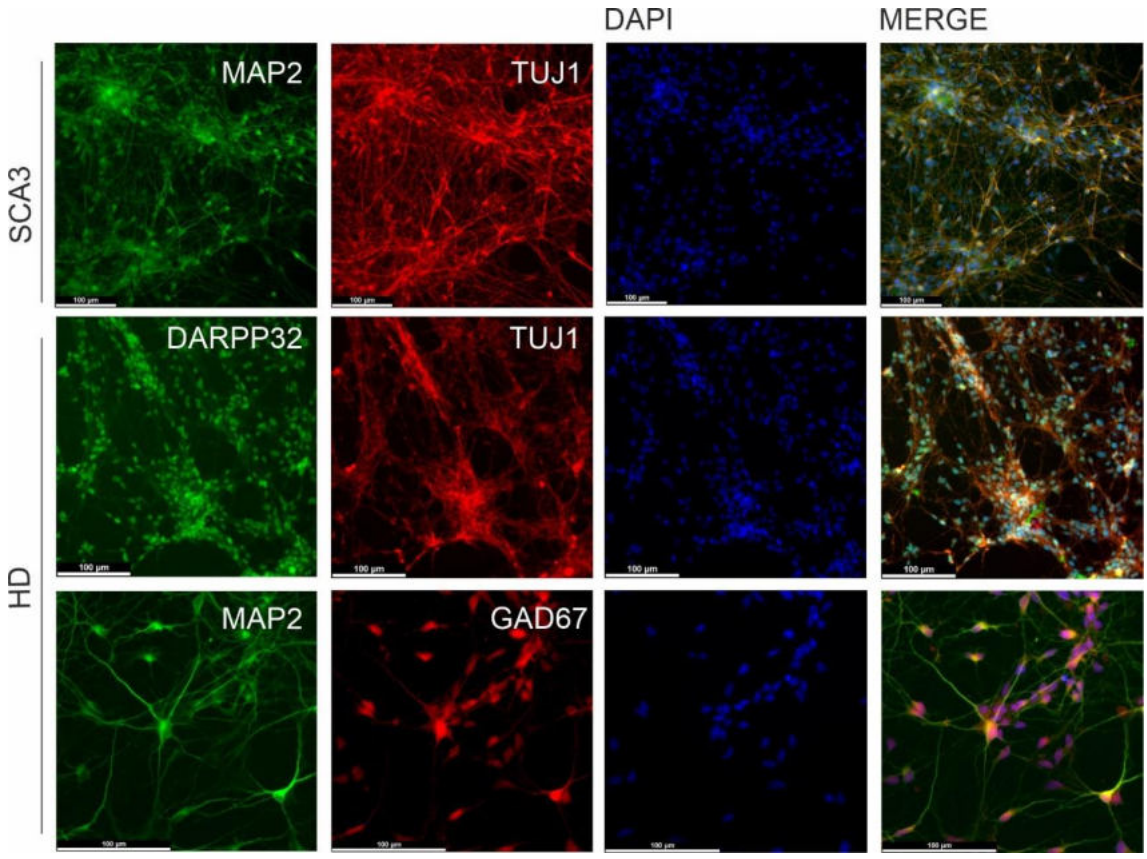
53. Minshall N, Git A. Enzyme- and gene-specific biases in reverse transcription of RNA raise concerns for evaluating gene expression. Sci Rep. 2020;10:8151.
54. Fodale V, Pintauro R, Daldin M, Altobelli R, Spiezia MC, Bisbocci M, et al. Analysis of mutant and total huntingtin expression in Huntington's disease murine models. Sci Rep. 2020;10:22137.
55. Liu JP, Zeitlin SO. Is huntingtin dispensable in the adult brain? J Huntington's Dis. 2017;6:1–17.
56. Saudou F, Humbert S. The biology of huntingtin. Neuron. 2016;89:910–26.
57. Tabrizi SJ, Leavitt BR, Landwehrmeyer GB, Wild EJ, Saft C, Barker RA, et al. Targeting huntingtin expression in patients with Huntington's disease. N Engl J Med. 2019;380:2307–16.
58. Wild EJ, Tabrizi SJ. Therapies targeting DNA and RNA in Huntington's disease. Lancet Neurol. 2017;16:837–47.
59. Wild EJ, Boggio R, Langbehn D, Robertson N, Haider S, Miller JRC, et al. Quantification of mutant huntingtin protein in cerebrospinal fluid from Huntington's disease patients. J Clin Invest. 2015;125:1979–86.
60. Macdonald D, Tessari MA, Boogaard I, Smith M, Pulli K, Szynol A, et al. Quantification assays for total and polyglutamine-expanded huntingtin proteins. PLoS One. 2014;9:e96854.
61. Southwell AL, Smith SE, Davis TR, Caron NS, Villanueva EB, Xie Y, et al. Ultrasensitive measurement of huntingtin protein in cerebrospinal fluid demonstrates increase with Huntington disease stage and decrease following brain huntingtin suppression. Sci Rep. 2015;5:12166.
62. Corey-Bloom J, Haque AS, Park S, Nathan AS, Baker RW, Thomas EA. Salivary levels of total huntingtin are elevated in Huntington's disease patients. Sci Rep. 2018;8:7371.
63. Caron NS, Banos R, Aly AE, Xie Y, Ko S, Potluri N, et al. Cerebrospinal fluid mutant huntingtin is a biomarker for huntingtin lowering in the striatum of Huntington disease mice. Neurobiol Dis. 2022;166:105652.
64. Ciolak A, Krzyzosiak WJ, Kozlowska E, Fiszer A. Generation of human iPS cell line IBCHi002-A from spinocerebellar ataxia type 3/Machado-Joseph disease patient's fibroblasts. Stem Cell Res. 2020;45:101796.
65. Hansen SK, Stummann TC, Borland H, Hasholt LF, Tümer Z, Nielsen JE, et al. Induced pluripotent stem cell - derived neurons for the study of spinocerebellar ataxia type 3. Stem Cell Res. 2016;17:306–17.
66. Fjodorova M, Li M. Robust induction of DARPP32-expressing GABAergic striatal neurons from human pluripotent stem cells. Huntington's Dis Methods Mol Biol. 2018;1780:585–605.
67. Whale AS, De Spiegelaere W, Trypsteen W, Nour AA, Bae Y-K, Benes V, et al. The Digital MIQE Guidelines Update: minimum information for publication of quantitative digital PCR experiments for 2020. Clin Chem. 2020;66:1012–29.

Publisher's Note



Additional file 1: Fig. S1.

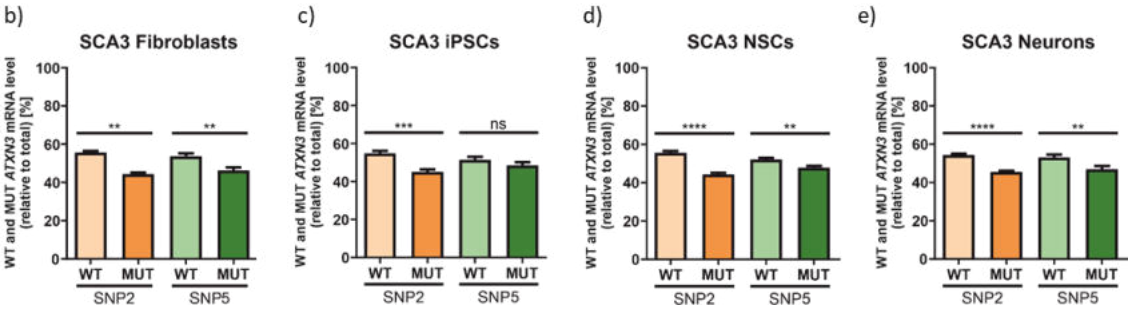
Analyses of *ATXN3* and *HTT* assays specificity. Results from ddPCR assays performed on seven samples with predefined ratios of WT/MUT *ATXN3* (A) and *HTT* (B-C) plasmids (I—100% WT and 0% MUT; II—90% WT and 10% MUT; III—75% WT and 25% MUT; IV—50% WT and 50% MUT; V—25% WT and 75% MUT; VI—10% WT and 90% MUT; VII—0% WT and 100% MUT) using ATXN3_SNP5 (A), HTT_SNP5 (B) and HTT_SNP7 assays (C). Mean/Precise values are indicated on WT/MUT bars ± poisson error.



Additional file 2: Fig. S2.
Immunocytochemistry performed on SCA3 and HD neurons. Exemplary images of immunofluorescent staining of SCA3 neurons for MAP 2 and TUJ1, and HD neurons for DARPP32, TUJ1, MAP 2 and GAD67. DAPI was used for nuclei staining.

a)

	SNP2		SNP5	
	%WT	%MUT	%WT	%MUT
SCA3 Fibroblasts	55.6	44.4	53.7	46.3
SCA3 iPSCs	54.9	45.1	51.4	48.6
SCA3 NSCs	55.7	44.3	52.1	47.9
SCA3 Neurons	54.5	45.5	53.0	47.0

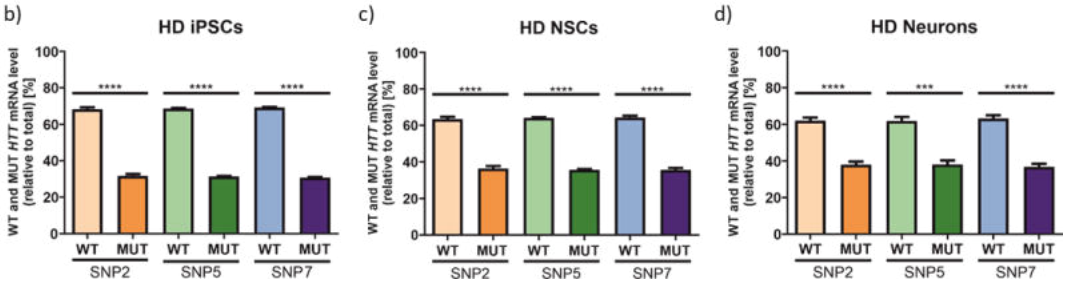


Additional file 3: Fig. S3.

Precise values describing WT/MUT ratio of *ATXN3* in selected SCA3 cells. A) Table with mean %WT and %MUT values obtained using four types of SCA3 cell lines and two SNP assays (mean values from three biological replicates). B-E) Results from ddPCR presented as a mean WT/MUT *ATXN3* transcript allele abundance obtained using ATXN3_SNP2 and ATXN3_SNP5 assays in fibroblasts (B), iPSCs (C), NSCs (D) and neurons (E). To determine statistical significance of differences between WT and MUT allele abundance data were analyzed using unpaired t test. For all experiments presented in this figure n=3. Two-tailed p value < 0.05 was considered significant and is depicted in the figure by: *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001. All data are presented as means ± SD.

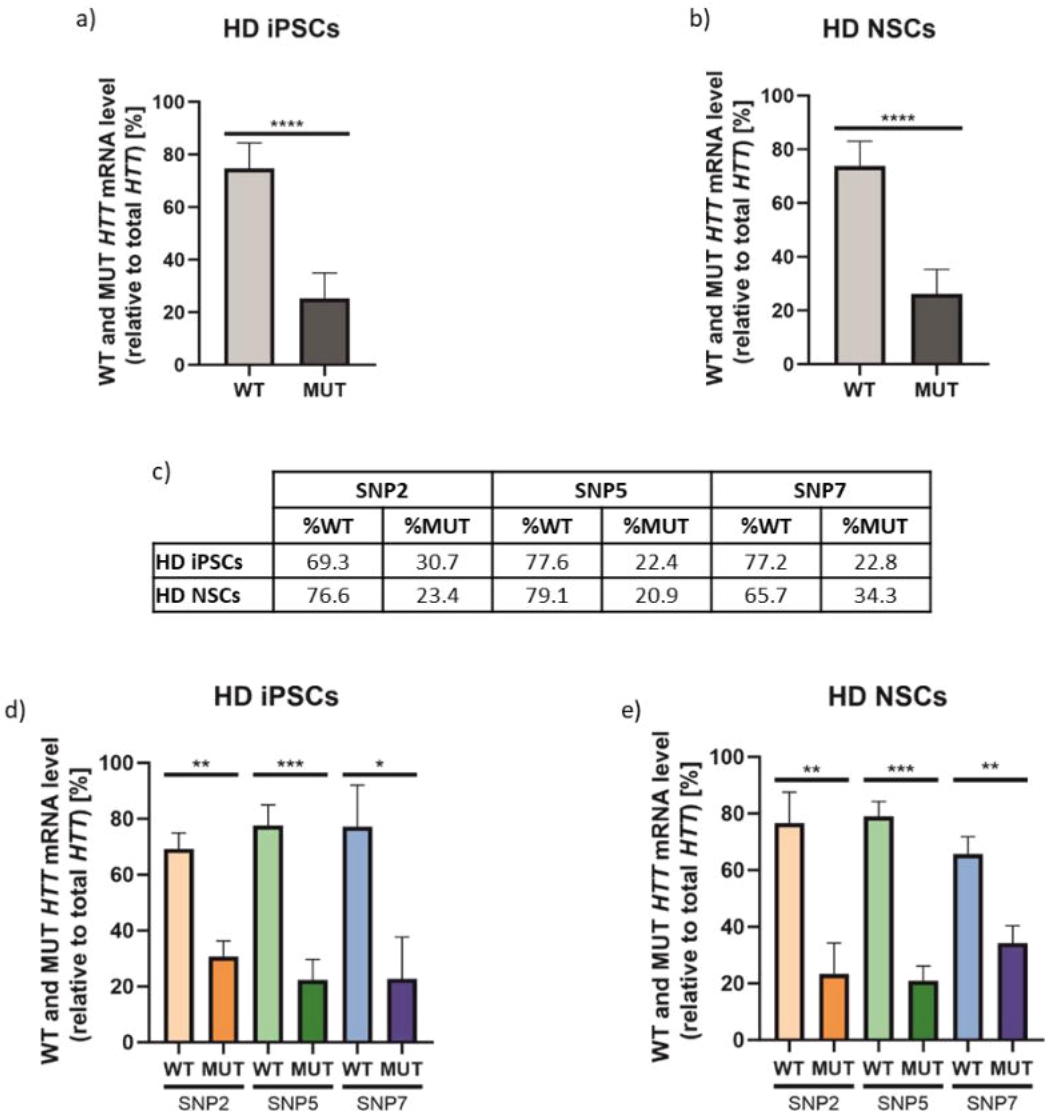
a)

	SNP2		SNP5		SNP7	
	%WT	%MUT	%WT	%MUT	%WT	%MUT
HD iPSCs	68.3	31.7	68.7	31.3	69.2	30.8
HD NSCs	63.5	36.5	64.2	35.8	64.3	35.7
HD Neurons	62.1	38.0	61.9	38.1	63.2	36.8



Additional file 4: Fig. S4.

Precise values describing WT/MUT ratio of *HTT* in selected HD cells. A) Table with mean %WT and %MUT values obtained using three types of HD cell lines and three SNP assays (mean values from three biological replicates). B-D) Results from ddPCR presented as a mean WT and MUT *HTT* transcript allele abundance obtained using HTT_SNP2, HTT_SNP5 and HTT_SNP7 assays in iPSCs (B), NSCs (C) and neurons (D). To determine statistical significance of differences between WT and MUT allele abundance, data were analyzed using unpaired t test. For all experiments presented in this figure n=3. Two-tailed p value < 0.05 was considered significant and is depicted in the figure by: *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001. All data are presented as means ± SD.

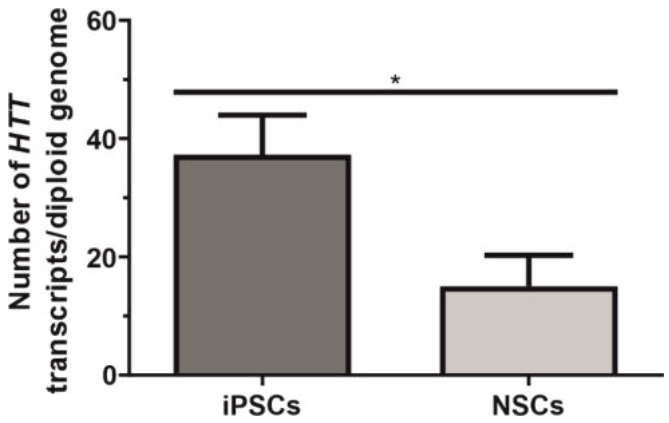


c)

	SNP2		SNP5		SNP7	
	%WT	%MUT	%WT	%MUT	%WT	%MUT
HD iPSCs	69.3	30.7	77.6	22.4	77.2	22.8
HD NSCs	76.6	23.4	79.1	20.9	65.7	34.3

Additional file 5: Fig. S5.

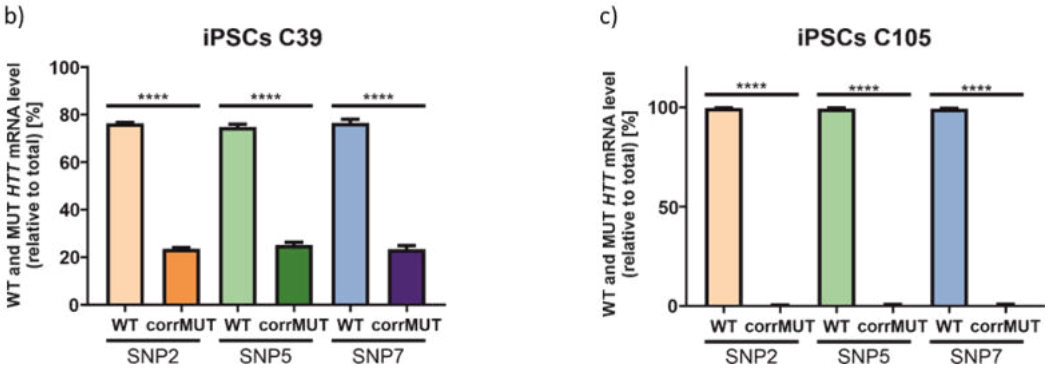
RNA-Seq analysis of *HTT* reads performed on HD iPSCs and NSCs. A-B) Results from RNA-Seq are presented as a mean WT and MUT *HTT* transcript allele abundance, calculated based on reads from three *HTT* SNP regions (SNP2, SNP5 and SNP7), in iPSCs (A) and NSCs (B). These data were analyzed using unpaired t test. C) Table summarizing precise RNA-Seq results. D-E) Precise results from RNA-Seq presented as a WT and MUT *HTT* transcript allele abundance based on reads from three, particular SNP regions (SNP2, SNP5 and SNP7) in iPSCs (D) and NSCs (E). Data were analyzed using unpaired t test. For all experiments presented in this figure *n*=3. All data are presented as means ± SD. Individual data values are available in Additional File 13.



Additional file 6: Fig. S6.
Estimation of the total number of endogenous *HTT* transcripts per diploid genome. Results were obtained using *HTT*_SNP7 assay and cDNA from HD iPSCs and NSCs. Data were analyzed using unpaired t test. For all experiments presented in this figure n=3. Two-tailed p value < 0.05 was considered significant and is depicted in the figure by: *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001. All data are presented as means ± SD.

a)

	SNP2		SNP5		SNP7	
	%WT	%corrMUT	%WT	%corrMUT	%WT	%corrMUT
iPSCs C39	76.4	23.6	74.8	25.2	76.5	23.5
iPSCs C105	99.7	0.3	99.5	0.5	99.3	0.7
iPSCs C7522	N/A	N/A	55.6	44.4	N/A	N/A



Additional file 7: Fig. S7.

Precise values describing WT/MUT ratio of *HTT* in isogenic controls to HD cells. A) Table with mean %WT and %corrMUT values obtained for lines and presented SNP assays (mean values from three biological replicates). B-C) Results from ddPCR presented as a mean WT/MUT *HTT* transcript allele abundance obtained using HTT_SNP2, HTT_SNP5 and HTT_SNP7 assays in C39 (B) and C105 (C). To determine statistical significance of differences between WT and MUT allele abundance, data were analyzed using unpaired t test. For all experiments presented in this figure n=3. Two-tailed p value < 0.05 was considered significant and is depicted in the figure by: *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001. All data are presented as means ± SD. Error bars in the figure represent standard deviations.

a)

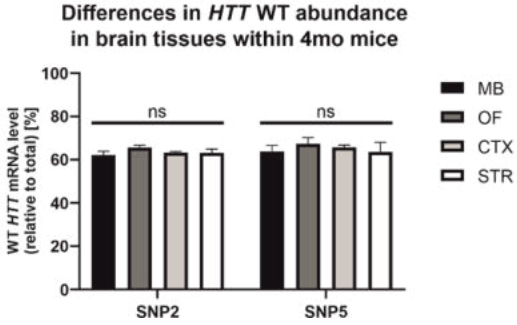
4mo Hu128/21 mice

	SNP2		SNP5	
	%WT	%MUT	%WT	%MUT
Midbrain (MB)	62.2	37.8	63.8	36.2
Olfactory bulb (OF)	65.7	34.3	67.4	32.6
Cortex (CTX)	63.3	36.7	65.7	34.3
Striatum (STR)	63.2	36.8	63.7	36.3

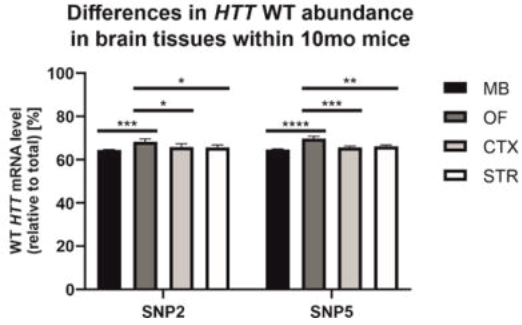
10mo Hu128/21 mice

	SNP2		SNP5	
	%WT	%MUT	%WT	%MUT
Midbrain (MB)	64.4	35.6	64.7	35.3
Olfactory bulb (OF)	68.3	31.7	69.8	30.2
Cortex (CTX)	65.9	34.1	65.6	34.4
Striatum (STR)	65.6	34.4	66.2	33.8

b)



c)



Additional file 8: Fig. S8.

Precise values describing WT/MUT ratio of *HTT* transgene in Hu128/21 mice. A) Table with mean %WT and %MUT values obtained using mice brain tissues and two ddPCR assays (mean values from three biological replicates). B-C) Results from ddPCR presented as a mean WT/MUT *HTT* transcript allele abundance obtained using HTT_SNP2 and HTT_SNP5 assays in 4-month-old (B) and 10-month-old mice (C). Data were analyzed using two-way ANOVA with Tukey's multiple comparison test. For all experiments presented in this figure n=3. Two-tailed p value < 0.05 was considered significant and is depicted in the figure by: *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001. All data are presented as means ± SD. Error bars in the figure represent standard deviations.

Primary antibodies	Dilution	Manufacturer
DARPP-32 (19A3)	1:400	Cell Signaling Technology (Cat. No. 2306)
GAD67	1:50	Santa Cruz Biotechnology (Cat. No. SC-28376)
MAP2	1:200	Cell Signaling Technology (Cat. No. 4542)
TUJ1	1:500	BioLegend (Cat. No. MMS-435P)
Secondary antibodies	**Dilution**	**Manufacturer**
anti-rabbit Alexa Fluor 488	1:1000	**Jackson ImmunoResearch** (711-546-152)
anti-mouse Alexa Fluor 594	1:1000	**Jackson ImmunoResearch** (715-586-151)

Supplementary Table 1. A list of primary and secondary antibodies with their dilutions used in immunocytochemistry.

Target Gene	Primer sequence (5' -> 3')	Description
ATXN3_Fwd	GGCCGTTGGCTCCAGAC	Cloning for SNP identification
ATXN3_Rev	AGATCCACTAAGTACTGTGACTTC	Cloning for SNP identification
HTT_Fwd	TTCTGCTTTTACCTGCGGCC	Cloning for SNP identification
HTT_Rev	AGTGTTCCCAAAGCCTGCTCAC	Cloning for SNP identification
HTT_RT	CTGTCTCAGAGCTGCTGACATAAC	RT for SNP identification
ATXN3_SNP2		ddPCR assay ID: dMDS306026999
ATXN3_SNP5		ddPCR assay ID: dHsaMDS942717091
HTT_SNP2		ddPCR assay ID: dMDS291328664
HTT_SNP5		ddPCR assay ID: dMDS735189637
HTT_SNP7		ddPCR assay ID: dMDS165698820
ACTB		ddPCR assay ID: dHsaCPE5190199
Rpp30		ddPCR assay ID: dMmuCNS822293939

Supplementary Table 2. Table listing all primers and ddPCR assays used in this research with their description.

Name	Sense (5'-3')	Antisense (5'-3')
A2	GCUGCUGC**A**GCUGCUGCUGCU	GCUGCUGC**A**GCUGCUGCUGCU
siHTT	GCCUUCGAGUCCCUCAAGUCC	ACUUGAGGGACUCGAAGGCCU
siRluc	AUCUGAAGAAGGAGAAAAATT	AUCUGAAGAAGGAGAAAAATT

Supplementary Table 3. Sequences of oligonucleotides used for silencing of endogenous HTT in HD NSCs.

Additional file 13. Individual data values for selected figures.

Determination of endogenous WT and MUT ATXN3 alleles expression ratios in SCA3 cell lines

		SNP 2				SNP 5			
		WT	MUT	%WT	%MUT	WT	MUT	%WT	%MUT
Fibroblasts	R1	142	114	55,47	44,53	65	58,6	52,59	47,41
	R2	125	102	55,07	44,93	92	74	55,42	44,58
	R3	156	122	56,12	43,88	115	102	53,00	47,00
	Mean			**55,55**	**44,45**			**53,67**	**46,33**
	Std. Dev.			**0,74**	**0,74**			**1,53**	**1,53**
iPSCs	R1	188	149	55,79	44,21	91	80	53,22	46,78
	R2	202	163	55,34	44,66	99	98	50,25	49,75
	R3	146	127	53,48	46,52	55,8	54	50,82	49,18
	Mean			**54,87**	**45,13**			**51,43**	**48,57**
	Std. Dev.			**1,22**	**1,22**			**1,57**	**1,57**
NSCs	R1	256	196	56,64	43,36	187	171	52,23	47,77
	R2	225	181	55,42	44,58	165	147	52,88	47,12
	R3	335	275	54,92	45,08	230	219	51,22	48,78
	Mean			**55,66**	**44,34**			**52,11**	**47,89**
	Std. Dev.			**0,88**	**0,88**			**0,84**	**0,84**
Neurons	R1	210	179	53,98	46,02	177	156	53,15	46,85
	R2	225	189	54,35	45,65	163	154	51,42	48,58
	R3	200	163	55,10	44,90	171	143	54,46	45,54
	Mean			**54,48**	**45,52**			**53,01**	**46,99**
	Std. Dev.			**0,57**	**0,57**			**1,52**	**1,52**

Determination of endogenous WT and MUT HTT alleles expression ratios in HD cell lines

		SNP 2				SNP 5				SNP 7			
		WT	MUT	%WT	%MUT	WT	MUT	%WT	%MUT	WT	MUT	%WT	%MUT
iPSCs	R1	551	266	67,44	32,56	309	140	68,82	31,18	291	131	68,96	31,04
	R2	456	215	67,96	32,04	243	110	68,84	31,16	250	110	69,44	30,56
	R3	526	232	69,39	30,61	293	136	68,30	31,70	294	131	69,18	30,82
	Mean			**68,26**	**31,74**			**68,65**	**31,35**			**69,19**	**30,81**
	Std. Dev.			**1,01**	**1,01**			**0,31**	**0,31**			**0,24**	**0,24**
NSCs	R1	300	175	63,16	36,84	233	132	63,84	36,16	210	121	63,44	36,56
	R2	269	161	62,56	37,44	216	120	64,29	35,71	197	104	65,45	34,55
	R3	277	150	64,87	35,13	201	111	64,42	35,58	177	99	64,13	35,87
	Mean			**63,53**	**36,47**			**64,18**	**35,82**			**64,34**	**35,66**
	Std. Dev.			**1,20**	**1,20**			**0,31**	**0,31**			**1,02**	**1,02**
Neurons	R1	202	114	63,92	36,08	155	86	64,32	35,68	136	72,8	65,13	34,87
	R2	190	119	61,49	38,51	122	80,5	60,25	39,75	117	72,2	61,84	38,16
	R3	229	148	60,74	39,26	144	92	61,02	38,98	141	84	62,67	37,33
	Mean			**62,05**	**37,95**			**61,86**	**38,14**			**63,21**	**36,79**
	Std. Dev.			**1,66**	**1,66**			**2,16**	**2,16**			**1,71**	**1,71**

Determination of endogenous HTT alleles ratios in genetically corrected HD iPSCs and non-HD iPSCs

		SNP 2				SNP 5				SNP 7			
		WT	**MUT**	**%WT**	**%MUT**	**WT**	**MUT**	**%WT**	**%MUT**	**WT**	**MUT**	**%WT**	**%MUT**
iPSCs C39	R1	329	101	76,51	23,49	205	73,6	73,58	26,42	190	62	75,40	24,60
	R2	379	120	75,95	24,05	239	79,5	75,04	24,96	235	74,5	75,93	24,07
	R3	365	111	76,68	23,32	239	76,3	75,80	24,20	225	62,6	78,23	21,77
	Mean			**76,38**	**23,62**			**74,81**	**25,19**			**76,52**	**23,48**
	Std. Dev.			**0,38**	**0,38**			**1,13**	**1,13**			**1,51**	**1,51**
iPSCs C105	R1	313	1	99,68	0,32	193	0,9	99,54	0,46	175	1,4	99,21	0,79
	R2	212	0,9	99,58	0,42	141	1,1	99,23	0,77	135	1	99,26	0,74
	R3	247	0,52	99,79	0,21	152	0,63	99,59	0,41	151	0,73	99,52	0,48
	Mean			**99,68**	**0,32**			**99,45**	**0,55**			**99,33**	**0,67**
	Std. Dev.			**0,11**	**0,11**			**0,20**	**0,20**			**0,17**	**0,17**
C7522	R1					214	174	55,15	44,85				
	R2					194	160	54,80	45,20				
	R3					214	162	56,91	43,09				
	Mean							**55,62**	**44,38**				
	Std. Dev.							**1,13**	**1,13**				

Silencing of endogenous HTT expression in HD NSCs

		siRL				A2				siHTT			
		WT	**MUT**	**%WT**	**%MUT**	**WT**	**MUT**	**%WT**	**%MUT**	**WT**	**MUT**	**%WT**	**%MUT**
R1	SNP2	133	65,4	67,04	32,96	104	37,5	73,50	26,50	66	33,9	66,07	33,93
	SNP5	111	50	68,94	31,06	90	34,7	72,17	27,83	61	31,7	65,80	34,20
	SNP7	99	46,8	67,90	32,10	84	31,1	72,98	27,02	52,7	23,4	69,25	30,75
	Mean			**67,96**	**32,04**			**72,88**	**27,12**			**67,04**	**32,96**
	Std. Dev.			**0,96**	**0,96**			**0,67**	**0,67**			**1,92**	**1,92**
R2	SNP2	165	76	68,46	31,54	147	56	72,41	27,59	68,8	33,3	67,38	32,62
	SNP5	116	57	67,05	32,95	113	46	71,07	28,93	60,6	31,5	65,80	34,20
	SNP7	105	51	67,31	32,69	102	39	72,34	27,66	47,7	24,5	66,07	33,93
	Mean			**67,61**	**32,39**			**71,94**	**28,06**			**66,42**	**33,58**
	Std. Dev.			**0,75**	**0,75**			**0,76**	**0,76**			**0,85**	**0,85**
R3	SNP2	186	83	69,14	30,86	126	47,7	72,54	27,46	77	40,9	65,31	34,69
	SNP5	127	65,2	66,08	33,92	94	36,6	71,98	28,02	64	34,6	64,91	35,09
	SNP7	115	56,1	67,21	32,79	83,5	30,8	73,05	26,95	58	28,4	67,13	32,87
	Mean			**67,48**	**32,52**			**72,52**	**27,48**			**65,78**	**34,22**
	Std. Dev.			**1,55**	**1,55**			**0,54**	**0,54**			**1,18**	**1,18**

4 months old mice

		SNP 2				SNP 5			
		WT (BAC)	MUT (YAC)	%WT	%MUT	WT (BAC)	MUT (YAC)	%WT	%MUT
MB	Mouse 1	587	330	64,01	35,99	571	295	65,94	34,06
	Mouse 2	632	391	61,78	38,22	586	317	64,89	35,11
	Mouse 3	194	125	60,82	39,18	45,2	29,4	60,59	39,41
	Mean			**62,20**	**37,80**			**63,81**	**36,19**
	Std. Dev.			**1,64**	**1,64**			**2,83**	**2,83**
OF	Mouse 1	1186	587	66,89	33,11	1157	497	69,95	30,05
	Mouse 2	1117	601	65,02	34,98	916	436	67,75	32,25
	Mouse 3	229	123	65,06	34,94	59,4	32,9	64,36	35,64
	Mean			**65,66**	**34,34**			**67,35**	**32,65**
	Std. Dev.			**1,07**	**1,07**			**2,82**	**2,82**
CTX	Mouse 1	1013	574	63,83	36,17	920	455	66,91	33,09
	Mouse 2	841	494	63,00	37,00	764	413	64,91	35,09
	Mouse 3	125	73	63,13	36,87	41,6	22,1	65,31	34,69
	Mean			**63,32**	**36,68**			**65,71**	**34,29**
	Std. Dev.			**0,45**	**0,45**			**1,06**	**1,06**
STR	Mouse 1	1117	606	64,83	35,17	947	468	66,93	33,07
	Mouse 2	1002	578	63,42	36,58	843	446	65,40	34,60
	Mouse 3	171	108	61,29	38,71	40	28,2	58,65	41,35
	Mean			**63,18**	**36,82**			**63,66**	**36,34**
	Std. Dev.			**1,78**	**1,78**			**4,40**	**4,40**

10 months old mice

		SNP 2				SNP 5			
		WT (BAC)	MUT (YAC)	%WT	%MUT	WT (BAC)	MUT (YAC)	%WT	%MUT
MB	Hu 109	2650	1445	64,71	35,29	1925	1040	64,92	35,08
	Hu 608	808	447	64,38	35,62	685	379	64,38	35,62
	Hu 609	1097	611	64,23	35,77	871	473	64,81	35,19
	Średnia			**64,44**	**35,56**			**64,70**	**35,30**
	Std. Dev.			**0,25**	**0,25**			**0,29**	**0,29**
OF	Hu 109	2810	1231	69,54	30,46	2225	916	70,84	29,16
	Hu 608	1087	537	66,93	33,07	966	434	69,00	31,00
	Hu 609	1356	627	68,38	31,62	1183	516	69,63	30,37
	Średnia			**68,28**	**31,72**			**69,82**	**30,18**
	Std. Dev.			**1,30**	**1,30**			**0,93**	**0,93**
CTX	Hu 109	4200	2025	67,47	32,53	1859	998	65,07	34,93
	Hu 608	1320	707	65,12	34,88	1103	584	65,38	34,62
	Hu 609	1192	638	65,14	34,86	917	464	66,40	33,60
	Średnia			**65,91**	**34,09**			**65,62**	**34,38**
	Std. Dev.			**1,35**	**1,35**			**0,70**	**0,70**
STR	Hu 109	3760	1871	66,77	33,23	3550	1864	65,57	34,43
	Hu 608	1027	568	64,39	35,61	806	405	66,56	33,44
	Hu 609	1158	607	65,61	34,39	892	448	66,57	33,43
	Średnia			**65,59**	**34,41**			**66,23**	**33,77**
	Std. Dev.			**1,19**	**1,19**			**0,57**	**0,57**

RNA-Seq analysis of HTT reads performed on HD iPSCs and NSCs

		SNP2		SNP5		SNP7	
		%WT	%MUT	%WT	%MUT	%WT	%MUT
iPSCs	R1	75,47	24,53	84	16	80	20
	R2	64,41	35,59	69,57	30,43	61,11	38,89
	R3	68,09	31,91	79,31	20,69	90,48	9,52
	Mean	**69,32**	**30,68**	**77,63**	**22,37**	**77,20**	**22,80**
	Std. Dev.	**5,63**	**5,63**	**7,36**	**7,36**	**14,88**	**14,88**
NSCs	R1	86,67	13,33	73,33	26,67	58,82	41,18
	R2	78,13	21,87	83,33	16,67	70,37	29,63
	R3	65,06	34,94	80,49	19,51	68	32
	Mean	**76,62**	**23,38**	**79,05**	**20,95**	**65,73**	**34,27**
	Std. Dev.	**10,88**	**10,88**	**5,15**	**5,15**	**6,10**	**6,10**