ARTICLE

Full sequence of mutant huntingtin 3′-untranslated region and modulation of its gene regulatory activity by endogenous microRNA

Kyung-Hee Kim[1,2] · Kawther Abu Elneel[1] · Jun Wan Shin[1,2] · Jae Whan Keum[1] · David Seong [1] · Seung Kwak[3] ·
Ramee Lee[3] · James F. Gusella[1,4,5] · Marcy E. MacDonald[1,2,5] · Ihn Sik Seong[1,2] · Jong-Min Lee [1,2,5]

Abstract
Huntington's disease (HD) is caused by an expanded CAG trinucleotide repeat in the first exon of the huntingtin gene (*HTT*). Since the entire course of the disease starts from this dominant gain-of-function mutation, lowering total or mutant huntingtin mRNA/protein has emerged as an appealing therapeutic strategy. We reasoned that endogenous mechanisms underlying *HTT* gene regulation may inform strategies to target the source of the disease. As part of our investigation to understand how the expression of *HTT* is controlled, we performed (1) complete sequencing analysis for mutant *HTT* 3′-UTR and (2) unbiased screening assays to identify naturally-occurring miRNAs that could lower the *HTT* mRNA levels. By sequencing HD families inheriting the major European mutant haplotype, we determined the full sequence of *HTT* 3′-UTRs of the most frequent mutant (i.e., hap.01) and normal (i.e., hap.08) haplotypes, revealing 5 sites with alternative alleles. In subsequent miRNA activity assays using the full-length hap.01 and hap.08 3′-UTR reporter vectors and follow-up validation experiments, hsa-miR-4324 and hsa-miR-4756-5p significantly reduced *HTT* 3′-UTR reporter activity and endogenous HTT protein levels. However, those miRNAs did not show strong haplotype-specific effects. Nevertheless, our data highlighting full sequences of *HTT* 3′-UTR haplotypes, effects of miRNAs on *HTT* levels, and potential interaction sites provide rationale and promising targets for total and mutant-specific *HTT* lowering intervention strategies using endogenous and artificial miRNAs, respectively.

These authors contributed equally: Kyung-Hee Kim, Kawther Abu Elneel

Supplementary information The online version of this article (https://doi.org/10.1038/s10038-019-0639-8) contains supplementary material, which is available to authorized users.

✉ Jong-Min Lee
jlee51@mgh.harvard.edu

[1] Molecular Neurogenetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA

[2] Department of Neurology, Harvard Medical School, Boston, MA 02115, USA

[3] CHDI Foundation, Princeton, NJ 08540, USA

[4] Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

[5] Medical and Population Genetics Program, The Broad Institute of M.I.T. and Harvard, Cambridge, MA 02142, USA

Introduction

An expanded CAG trinucleotide repeat (>35) in the huntingtin (*HTT*) causes Huntington's disease (HD) [1]; the size of the expanded CAG repeat shows strong inverse correlations with age at onset of symptoms and age at death [2, 3]. Since an expanded CAG repeat is the primary determinant of the disease, targeting mutant *HTT* mRNA has emerged as an appealing therapeutic strategy [4, 5]. For example, gene silencing approaches directly targeting CAG repeat in *HTT* mRNA showed therapeutic benefit with some allele specificity in model systems [6]. However, CAG repeat-targeting intervention strategies may not provide sufficient gene specificity because many other CAG repeat-containing genes can be also affected [7]. Alternatively, silencing approaches targeting the genetic variations associated with the haplotypes carrying the disease-causing mutation may provide better gene specificity while achieving mutant allele specificity.

Although all cases of HD are due the same mutation, the causative mutation of HD sits on diverse haplotype

backbones [8–10]. For example, to account for approximately 90% of mutant chromosomes in HD subjects with European ancestry, 16 different haplotypes are required [9, 11–15]. Importantly, frequencies of *HTT* gene haplotypes in the HD mutant chromosomes are quite different from those in normal chromosomes [9, 10, 16]. For example, the most frequent mutant haplotype in European HD subjects (hap.01) is infrequent in normal chromosomes. Conversely, the most common normal haplotype in Europeans (hap.08) is relatively infrequent in HD mutant chromosomes. As a result, more than 90% of HD subjects have mutant haplotypes that are different from the normal counterparts [9]. These are particularly relevant with respect to therapeutics, because different alleles on mutant and normal haplotypes in a given HD subject are targets of mutant allele-specific gene targeting strategies, making investigation of *HTT* haplotypes and identification of associated alleles [8–10, 16, 17] highly significant.

Most widely used *HTT* targeting strategies rely on exogenous silencing reagents such as siRNA, ASO (antisense oligonucleotide), or CRISPR/Cas9 [4, 5, 7, 17–19]. Although such exogenous nucleotide-based reagents showed some pre-clinical efficacy in model systems of HD [20, 21], efficient delivery to relevant region remains as a significant challenge. Alternatively, we envisioned that therapeutic *HTT* lowering can be achieved endogenously. A potential mechanism that may be involved in natural *HTT* gene regulation and thus can be used to lower *HTT* mRNA expression levels is microRNA (miRNA). While others demonstrated the value of miRNAs as biomarkers of HD and tools to probe underlying disease mechanisms [22–26], we are primarily interested in understanding miRNA-mediated regulation of the *HTT* expression levels because of its therapeutic potential. Therefore, we set out to identify endogenous miRNAs that can lower the *HTT* mRNA levels and evaluated their mutant allele-specificities using full 3'-UTR sequences of the most common mutant and normal *HTT* haplotypes.

Materials and methods

Haplotype analysis to determine the full *HTT* 3′-UTR sequence

Reconstruction of the full haplotype sequences of *HTT* 3′-UTR was based on the whole genome sequencing data of selected HD families inheriting mutant hap.01 haplotype [12]. Our genetic analysis focused on 3′-UTR region of RefSeq transcript NM_002111 (chr4:3241786-3245687; GRCh37/hg19 coordinates; Fig. S1) in order to reconstruct the full sequence of *HTT* 3'-UTR of the most common mutant (hap.01) and normal haplotype (hap.08) in HD

subjects with European ancestry [9, 12, 16]. In both families, sequences of mutant hap.01 *HTT* 3'-UTR were fully reconstructed without any missing, generating complete 3'-UTR sequences. The normal hap.08 3'-UTR sequence was reconstructed based on un-transmitted chromosomes in trios #4 and #5 (Fig. S2). Allele frequencies of those 6 variations in the control population were based on 1000 Genomes Project data (phase 3, Europeans).

1000 Genomes Project data analysis to determine 8 common *HTT* haplotype 3′-UTR sequences

Sequences of 8 common *HTT* haplotype 3′-UTR were reconstructed from the 1000 Genomes Project data (phase3). Focusing on the most common 8 haplotypes, each phased haplotype was grouped based on our haplotype definitions [12, 16]. For a given haplotype comprising a group of chromosomes from 1000 Genomes Project participants, we identified the most frequent allele to identify representative alleles. Finally, sites with alternative alleles were identified for each haplotype to summarize sequences of 8 common haplotype 3'-UTR.

Cloning of the full *HTT* 3′-UTR into dual luciferase reporter vector

DNA sample from a HD subject carrying hap.01 mutant and hap.08 normal haplotypes (Coriell, GM01169; 17/44 CAG repeats) [9] was used to clone the full *HTT* 3′-UTR into a dual luciferase reporter vector. DNA sample was PCR amplified using a primer set (5′-TCCAGTCTGTGCTT GAGGTG-3′ and 5′-GCAACAACTGATGGTTTCCA-3′). Amplified DNA was digested (*XhoI* and *XbaI*), and inserted into a pmirGLO Dual-Luciferase reporter vector (Promega). The entire *HTT* 3'-UTR of hap.01 or hap.08 was independently inserted at the 3'-UTR site of *luc2* gene. Full sequences of resulting hap.01-pmirGLO and hap.08-pmirGLO reporter vectors were validated by Sanger sequencing.

Cell culture

The miRNA screening assays and validation experiments were performed using HEK293T cells and the HD patient-derived fibroblast. Cells were cultured in DMEM supplemented with 10% fetal bovine serum and Penicillin-Streptomycin at 37 °C and 5% CO_2.

The miRNA library and luciferase activity assays for the primary screening assay

The mirVana miRNA mimic library (1719 miRNAs) was purchased from the Life Technologies. Cells were

transfected using Lipofectamine 2000 (Invitrogen). Briefly, 20,000 cells/well (in 50 microliter) were plated in 96-well plates and transfected with 25 microliter of transfection mixture comprising 0.2 microliter/well of Lipofectamine, 30 nM of a test microRNA, and 80 ng/well of a reporter vector. All pipetting for screening assays were performed by an automated liquid handling system (Biomek FX, Beckman Coulter). We included (1) a control miRNA, hsa-125b-5p [25], (2) a miRNA negative control (scramble), (3) a positive control siRNA targeting *HTT* 3′-UTR (5′-CGTTGTACATGTTCCTGTT-3′), and (4) a negative control for siRNA (scramble). 24 h post-transfection, luciferase activity assays were performed using Dual-Glo luciferase reagent (Promega). Firefly luciferase reagent was added by the automated pump (50 microliter), plates were incubated for 15 min at room temperature, and then each well was read for 2 s by MicroLumatPlus LV 96 V (Berthold). 45 min after, 50 microliter of stop/*Renilla* solution was added, incubated for 15 min at room temperature, and each well was read for 2 s. Each miRNA was tested in duplicates in separate plates, and the entire screening assays were repeated twice.

Primary screening data analysis and identification of candidate miRNAs for follow-up

In our primary screening assays, each miRNA was tested 4 times. In addition, each test plate included wells for background (cells without transfection), basal reporter activity (cells transfected with reporter vector without miRNA treatment), miRNA negative control, miRNA control, siRNA negative control, and siRNA positive control for background correction and data normalization. Plate-specific background correction and normalization was performed to generate 4 data values (percent of control) for each test miRNA. Briefly, for a given plate, background activity of Firefly luciferase and *Renilla* luciferase were calculated by taking median values of background wells. These background values (mean of two) were subtracted from Firefly and *Renilla* luciferase values for miRNA negative control. Similarly, Firefly and *Renilla* luciferase values of test miRNAs were subtracted by the same background value. Finally, background corrected Firefly luciferase/*Renilla* luciferase ratio of a test miRNA was divided by Firefly luciferase/*Renilla* luciferase ratio of miRNA negative control to determine the effects of a test miRNA on *HTT* 3′-UTR (% of control).

Secondary assays, dose-response experiments, and validation

The primary screening assay results revealed 74 candidate miRNAs for secondary assays. We performed secondary reporter activity assays for those using hap.01, hap.08 and empty vector control dual luciferase reporters to determine effects on different haplotype 3′-UTR. Experimental conditions and data analysis for the secondary assay were same as those in the primary screening assay. Each candidate miRNA was tested in duplicate for the secondary activity assay. Subsequently, we identified miRNA that showed minimal effects on empty vector control but robust effects on hap.01 or hap.08 reporter vector to eliminate false positives. Four miRNAs survived secondary activity assays, and subjected to dose-response reporter activity assays. Potential targets of candidate miRNAs were predicted using the TargetScan program available in miRBase (http://www.mirbase.org/).

Statistical analysis

Data were presented as mean ± standard deviation. Statistical analyses were performed using GraphPad InStat program (student *t*-test). Also, multiple comparisons were conducted using Tukey's range test and Dunett's multiple comparison test. $^{*,\#}P < 0.05$; $^{**,\#\#}P < 0.01$; $^{***,\#\#\#}P < 0.001$.

Results

Expression patterns of *HTT* 3′-UTR in human tissues

Firstly, we characterized the expression patterns of *HTT* 3′-UTR in human tissues to select a region for sequencing and molecular analyses. Pioneering works revealed that differential 3′-polyadenylation of *HTT* generates two mRNA species with different sizes [27, 28]. Different transcripts appeared to have same coding sequence with different 3′-UTRs; a larger transcript with a long 3′-UTR (NM_002111; ~3.9KB) and a smaller transcript with a short 3′-UTR (L12392; ~0.6KB). Expression of these transcripts varies by tissues and developmental stages [27, 28]. In order to characterize expression patterns of 3′-UTR of *HTT* in adult human tissues at a high resolution, we examined the Human BodyMap 2.0 (http://www.ensembl.info/blog/2011/05/24/human-bodymap-2-0-data-from-illumina/), which surveyed 16 tissues using the Illumina HiSeq 2000 RNAseq platform. We utilized sequence coverage data to determine the *HTT* 3′-UTR region expressed in each tissue. Generally, RNAseq sequence reads were aligned over the entire long 3′-UTR region in all tissues examined (Fig. S1). However, certain tissues displayed an enrichment of sequencing reads at short 3′-UTR (skeletal muscle), 3′ end of long 3′-UTR (brain), or both (prostate). Differences in relative usage of the 3′-UTRs in different tissues (or cell types) are intriguing [29], suggesting cell type-specific regulatory mechanisms. Since the main scope of our study is to investigate miRNA-mediated

gene regulation of *HTT* applicable to brain, we selected the long 3′-UTR (chr4:3241786-3245687, GRCh37/hg19) for subsequent genetic and molecular analyses.

Full sequence of the *HTT* 3′-UTR major mutant haplotype

In the long *HTT* 3′-UTR region (chr4:3241786-3245687), 988 single nucleotide polymorphisms (SNPs) are annotated in the dbSNP (build 151), but their alleles in the *HTT* haplotypes are mostly unknown. We thus determined the full 3′-UTR sequence of the most frequent mutant (hap.01) and normal (hap.08) *HTT* haplotypes [12] (Fig. S2). Alleles of nine transmitted mutant hap.01 chromosomes were consistent at all sites, unequivocally generating the full hap.01 haplotype *HTT* 3′-UTR sequence. In addition, 2 children carrying two hap.01 haplotypes (Fig. S2; trio #7 and #9) did not show any heterozygous sites in the region, indicating that the 3′-UTR sequences are same between mutant hap.01 and normal hap.01 in those individuals. We also determined the normal hap.08 3'-UTR sequence, based on un-transmitted chromosomes in trio #4 and #5 (Fig. S2) (summarized in Fig. 1). For example, alleles of hap.01 (red) and hap.08 (green) are different from the reference sequence (black) at 3 and 4 locations, respectively. (Fig. 1a, blue triangles). Between mutant hap.01 and normal hap.08 haplotypes, alleles at 5 SNP sites were different (filled blue

triangles in Fig. 1a), providing potential target sites for allele-specific silencing strategies for HD subjects with the most common diplotype.

Next, we extended our genetic analysis to the 8 most common haplotypes [12, 16] in HD subjects. We determined the *HTT* 3′-UTR sequences of those haplotypes using publicly available sequence/phase data, revealing 8 sites where at least one the most common 8 haplotypes carries alternative allele (Fig. S3). Because CAG expansion mutations on existing normal haplotypes generated the HD mutant chromosomes [9, 11–14], the 3′-UTR sequences of normal and mutant haplotypes are hypothesized to be same for a given haplotype (Fig. S2).

Unbiased miRNA activity screening assays using the full-length hap.01 3′-UTR reporter vector

We next tested whether naturally-occurring miRNAs interact with the *HTT* 3′-UTR. For this, we generated pmirGLO-hap.01-3′-UTR (namely, hap.01 reporter vector; Fig. S4A, middle panel) and pmirGLO-hap.08-3′-UTR (namely, hap.08 reporter vector; Fig. S4A; right panel). These *HTT* 3′-UTR reporter vectors produce Firefly luciferase mRNA with respective *HTT* 3′-UTR and control *Renilla* luciferase (for data normalization). Thus, interaction between a miRNA and *HTT* 3′-UTR is predicted to induce translational suppression or mRNA destabilization [30, 31],
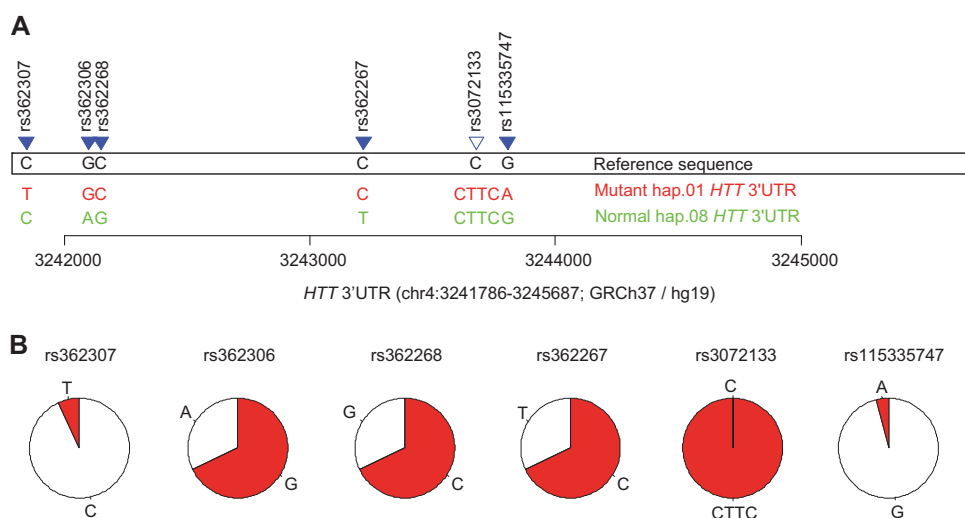

Fig. 1 The full *HTT* 3′-UTR sequences of the most common HD mutant and normal haplotypes. **a** In HD subjects with European ancestry, hap.01 is the most common HD disease haplotype. The most common normal haplotype in the control population and normal chromosomes in HD subjects is hap.08. As expected from the individual haplotype frequency, the most common diplotype in HD subjects is the combination of hap.01 (mutant) and hap.08 (normal). Focusing on those two haplotypes, full sequences of *HTT* 3′-UTR were determined through genetic phasing analysis of HD families carrying mutant hap.01 and normal hap.08 haplotypes. The hap.01 (red) and hap.08 (green) carry alternative alleles at 3 and 4 sites, respectively (blue triangles); two haplotypes have different alleles at 5 locations (filled blue triangles). Genomic coordinate was based on the GRCh37/hg19 assembly. **b** Allele frequencies of 6 SNPs were based on 1000 Genomes Project data (Phase 3; Europeans). In each pie chart, the frequency of the allele of the hap.01 haplotype is depicted in red. The frequency of alternative allele for SNP rs3072133 is 100% in 1000 Genomes Project data

leading to lower levels of Firefly luciferase activity without altering *Renilla* luciferase activity (Fig. S4B).

Using the hap.01 reporter vector, we performed primary screening assays in HEK293T cells to test a library of 1719 unique miRNAs (Fig. S5A). Each miRNA was tested in duplicate in two separate screening assays, generating 4 data points for each miRNA (Fig. S5B). Then, candidate miRNAs were identified for secondary luciferase activity assays using 3 reporter vectors (empty vector control, hap.01, and hap.08 reporter vectors) to judge their allele specificity (Fig. S5). In the primary screening assays, we included 1) a positive control siRNA and 2) a miRNA that has been reported to reduce *HTT* levels (hsa-miR-125b-5p) [25] (Fig. S6A). A positive control siRNA (Fig. S6B) and hsa-miR-125b-5p (Fig. S6C) reduced the Firefly luciferase activity of hap.01 reporter by 46 and 38%, respectively. A high level of reproducibility between replicates of test miRNAs was observed in the primary assay (Fig. S7), supporting the overall quality. Next, we performed statistical analysis of primary assay results of 1719 miRNAs (Table S1) to identify candidates that significantly reduced the hap.01 3′-UTR reporter activity. We applied the following filtering to identify candidates for secondary reporter activity assays: (1) effect size (percent of negative control < 85%), (2) variability (coefficient of variation < 20%) and (3) significance (false discovery rate < 0.05). These criteria identified 74 candidate miRNAs for the secondary activity assay (Fig. 2).

Secondary reporter activity assay and validation

We reasoned that the hit miRNAs based on reduced Firefly luciferase activity in the primary assays could be due to (1) an interaction between the test miRNA and 3′-UTR of the Firefly reporter, (2) an interaction between the test miRNA and Firefly luciferase mRNA coding sequence, or (3) toxicity. In order to eliminate false positives and to evaluate allele specificity, the 74 candidate miRNAs were further tested on (1) empty reporter vector, (2) the hap.01 reporter vector, and (3) the hap.08 reporter vector (Fig. S4) in secondary reporter activity assays (Fig. S8). We anticipated that a genuine miRNA that interacts exclusively with the *HTT* 3′-UTR would alter the activity of reporter vectors containing *HTT* 3′-UTR, not empty vector.

As shown in the secondary assays, some candidate miRNAs strongly reduced the activity of the empty vector control lacking the *HTT* 3′-UTR (Fig. S8A, circles at the bottom left corner), indicating that those miRNAs interact with Firefly luciferase sequence, representing false positives. In addition, some candidate miRNAs did not reduce the hap.01 reporter vector activity in the secondary assays (Fig. S8B), indicating that they were also false positives. To identify true positives, we identified miRNAs that showed minimal impact on the empty vector control (90–110% of negative control) but strongly influenced the *HTT* 3′-UTR vector assays (>30% reduction of *HTT* 3′-UTR vectors). This revealed 4 and 5 miRNAs that significantly reduced hap.01 (Fig. 3a) and
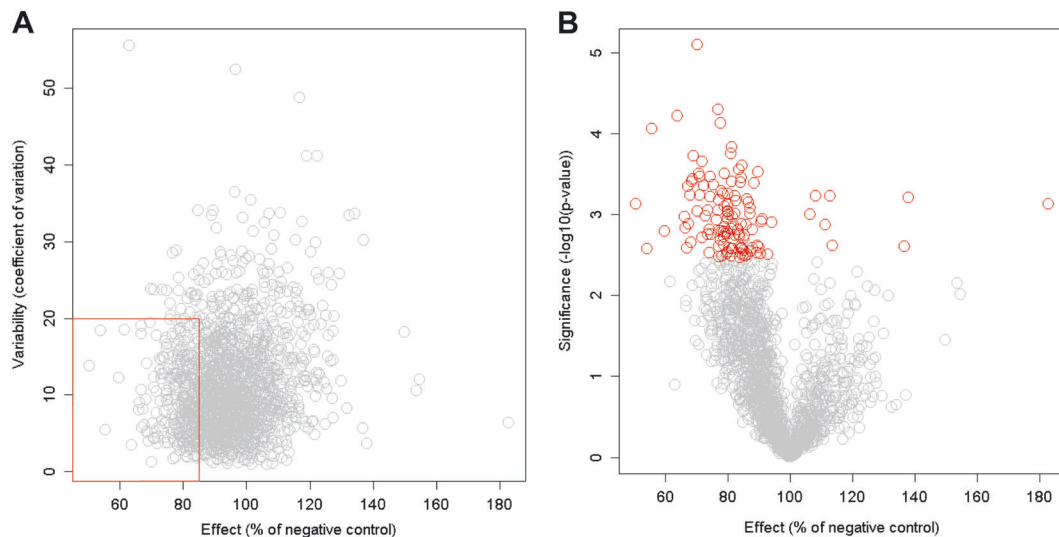

Fig. 2 Unbiased screening assays reveal candidate miRNAs that may interact with *HTT* 3′-UTR. Using a reporter vector containing the full length *HTT* 3′-UTR sequence of the hap.01 haplotype, we performed miRNA screening assays to evaluate a set of 1719 independent, well-annotated miRNAs. Each miRNA was assayed in duplicate plates, and the entire screening assays were repeated, generating 4 data points for a given miRNA. Starting from 1719 miRNAs, we identified 74 candidate miRNAs for secondary assays based on effect sizes (<85% of negative control; a red vertical line in panel A), relative variability (coefficient of variation < 20%; a red horizontal line in (**a**), and significance (false discovery rate < 0.05; red circles in (**b**). Each circle represents a miRNA. Y-axis in panel B represents nominal *p*-value, and miRNAs with false discovery rate < 0.05 were highlighted in red
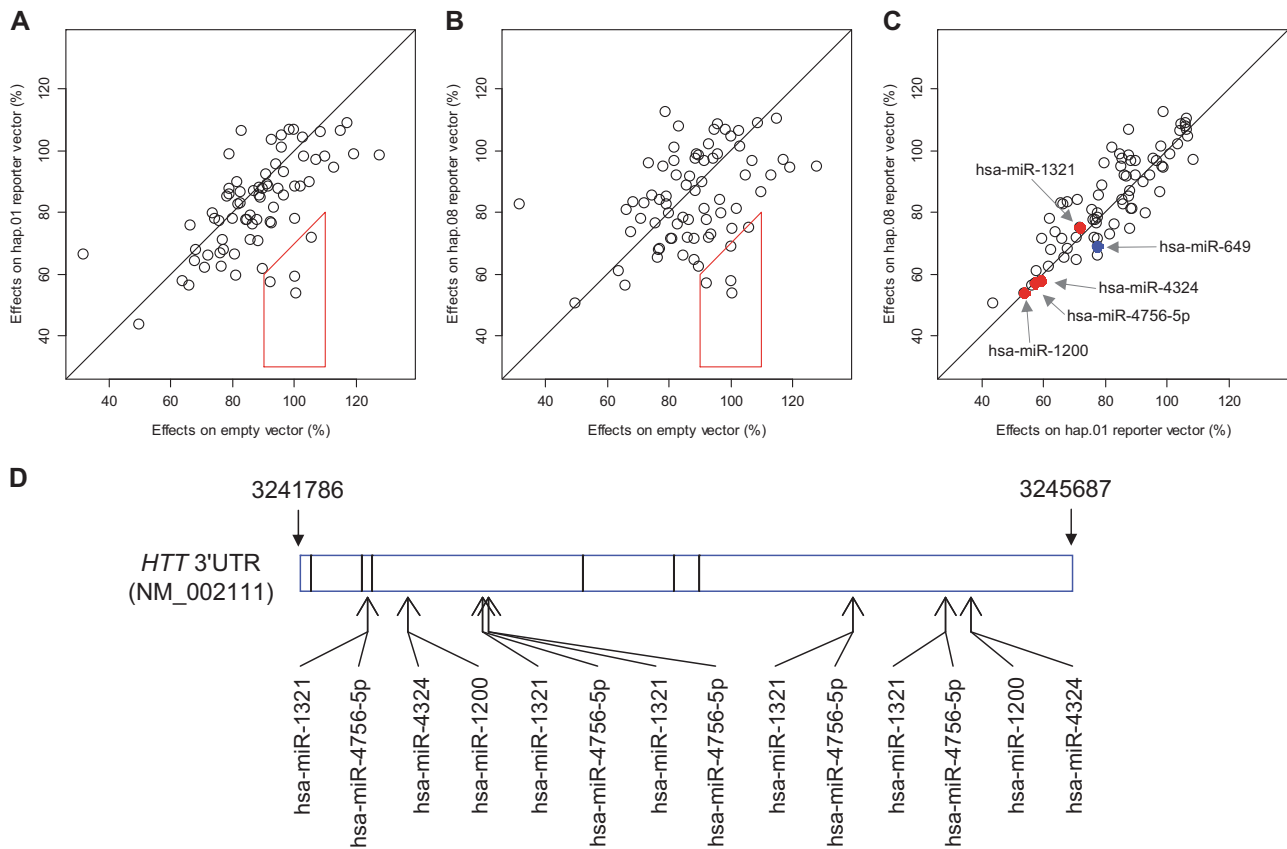
Fig. 3 *HTT* 3′-UTR-interacting miRNAs, and their predicted interaction sites. Secondary activity assays were performed using empty vector control, hap.01 3′-UTR reporter vector, and hap.08 3′-UTR reporter vector for 74 candidate miRNAs in order to identify miRNAs that truly interact with *HTT* 3′-UTR sequence. Effects of a miRNA were compared between (**a**) empty vector and hap.01-3′-UTR vector, (**b**) empty vector and hap.08-3′-UTR vector, and (**c**) hap.01-3′-UTR and hap.08-3′-UTR vectors. A group of miRNAs that are inside of red trapezoids showed reduction of 3′-UTR activity in either hap.01-3′-UTR vector (**a**) or hap.08-3′-UTR vector (**b**) without impacting empty vector control significantly. Four (**a**) and five miRNAs (**b**) significantly reduced the activity of *HTT* 3′-UTR reporter vector. Among those, 4 shared miRNAs reduced hap.01 and hap.08 reporter activities without impacting the empty vector. Allele specificity of each candidate miRNA was evaluated by comparing effects on hap.08 vector (**c**, Y-axis) to those on hap.01 vector (**c**, X-axis). Four positive hits are indicated in filled red circles (**c**). One miRNA (has-miR-649) showed slightly weaker effects on hap.01 reporter vector (a blue circle in **c**). Predicted interaction sites of those 4 miRNAs were shown relative to *HTT* 3′-UTR (**d**). Vertical lines in (**d**) marks locations of six variation sites described in Fig. 1

hap.08 3'-UTR reporter activity (Fig. 3b), respectively. Among them, 4 candidate miRNAs (hsa-miR-1200, hsa-miR-1321, hsa-miR-4324, and hsa-miR-4756-5p) met the criteria in both hap.01 and hap.08 reporter vector assays (Fig. 3c), reflecting non-haplotype-specific effects. These 4 shared miRNAs are predicted (www.targetscan.org/) to interact with *HTT* 3′-UTR sequence at various locations (Fig. 3d, vertical lines in the rectangle), revealing potential *HTT* 3′-UTR regions accessible to miRNAs. These potential interaction sites do not overlap with any of the genetic variations that distinguish hap.01 from hap.08 (Fig. 3d), consistent with the non-haplotype-specific effects of those 4 candidate miRNAs (Fig. 3a, red circles). The miRNA hsa-miR-649 reduced 22.4 and 31.4% of hap.01 and hap.08 reporter vector activity, respectively (Fig. 3c, blue circle), showing a modest hap.08-specificity. Therefore, this miRNA was not analyzed in the validation analysis.

Validation of candidates

We further determined (1) the dose-response relationship using the reporter vectors and (2) the impact on endogenous huntingtin protein (HTT) expression levels for the four shared candidate miRNAs. In reporter vector activity assays, 3 candidate miRNAs (i.e., hsa-miR-1200, hsa-miR-4324, hsa-miR-4756-5p) showed robust dose-response relationships in both hap.01 (open circles) and hap.08 reporter assays (filled triangles) without significantly altering empty vector control (filled circles) (Fig. S9). In addition, immunoblot analysis following miRNA treatment showed a notable reduction of total huntingtin levels for hsa-miR-1200, hsa-miR-4324, hsa-miR-4756-5p in HEK293T cells (Fig. 4a). However, hsa-miR-1200 also significantly reduced Actin protein levels. Since four candidate miRNAs did not significantly impact on overall cell
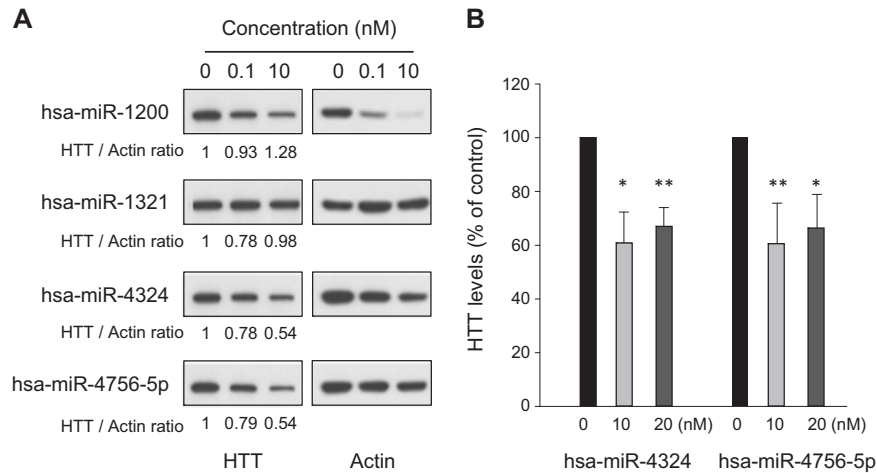
Fig. 4 Impacts of candidate miRNAs on the levels of endogenous HTT protein. Four candidate miRNAs from secondary analyses were further tested in HEK293T cells (a). Band intensity of HTT was normalized by levels of Actin, and then vehicle (PBS) treated sample was set to 1 to calculate HTT/Actin ratio. Western blot analysis was performed for each miRNA using MAB2166 antibody (Millipore) to detect both mutant and normal HTT protein. Subsequently, hsa-miR-4324 and hsa-miR-4756-5p were tested for viability on HD-derived fibroblasts (GM01169 in b). Data represent mean and standard errors of three independent experiments. $*P < 0.05$; $**P < 0.01$

viability (Fig. S10), reduced HTT and Actin protein levels by hsa-miR-1200 imply non-specific targeting or multi-targeting [32–34]. Considering targeting efficacy and gene specificity, hsa-miR-4324 and hsa-miR-4756-5p were further tested on a fibroblast line from a HD subject who carries hap.01 mutant and hap.08 normal haplotypes (GM01169) [9]. As shown in Fig. 4b and Fig. S11, those two miRNAs significantly reduced total HTT protein levels in HD patient-derived cells, showing consistent effects on reporter vectors, and endogenous *HTT* in non-HD cells and HD cells.

Discussion

In HD, miRNAs have been investigated to (1) understand pathogenesis [22, 24, 35–37], (2) identify disease biomarkers [26, 38, 39], (3) investigate *HTT* gene regulation [25, 40], and (4) evaluate pre-clinical therapeutic potential [38, 41–43]. By sequencing small RNAs in post-mortem brains, 5 miRNAs were founded to be elevated in HD brains (BA9 cortex), and ectopic expression of miR-10b-5p increased survival in a cell model of HD. Differentially expressed miRNAs in HD compared to normal controls may be useful in understanding underlying pathological mechanisms and potential biomarkers in intervention studies. Developing therapeutic huntingtin lowering strategies based on those may be therefore challenging as their differential expression in HD compared to normal controls may represent cellular compensatory responses [22].

Although miRNAs are useful for disease biomarkers [22–24, 26, 37–39, 44–46], their value as a therapeutic

reagent is highly significant considering the fact that one of the miRNAs' primary functions of miRNA is to regulate the expression levels of genes by binding to 3′-UTR or coding region of target mRNAs [47]. We reasoned that miRNA-mediated gene regulation mechanisms for *HTT*, if they exist, inform means to efficiently intervene in HD pathogenesis because the source of the disease can be addressed by endogenous molecules. Due to miRNAs' pivotal functions in various biological pathways, identification of targets is one of the most important research topics in the field. Despite its importance, systemic identification of targets of miRNAs has been hampered due to the lack of high-throughput experimental approaches. Since (1) Watson-Crick base pairing between ~7 nucleotide miRNA seed sequence and target mRNA is important for post-transcriptional gene repression [48] and (2) well-annotated genome sequences of many organisms are readily available, many computational algorithms to instead 'predict' targets of miRNAs have been developed [49, 50]. The abundance of readily available prediction algorithms has, however, generated situations where investigators chose target prediction programs based upon their own subjective preferences. Although target prediction algorithms have significantly contributed to the field, they have somewhat limited values because most prediction programs (1) are based on the reference genome sequence, which does not represent the real haplotypes/sequences in the population (Fig. S3), and (2) do not consider mRNA secondary structure. Those inherent limitations of target prediction approaches may generate false positive and false negative findings in a set of predicted targets of miRNAs. Since we are focused on a single gene (i.e., *HTT*), we reasoned that

unbiased experimental approaches testing ~1700 well-annotated miRNAs might be feasible. In addition, our unbiased approach would not involve target prediction algorithms and therefore subsequently identified *HTT*-interacting miRNAs might have fewer false-positives and false-negatives compared to prediction-based methods. To overcome previously mentioned two major limitations of prediction-based approaches, we constructed full length *HTT* 3′-UTR reporter vectors based on the most frequent mutant and normal haplotypes in HD subjects with European ancestry. Subsequently, we performed unbiased miRNA library screening assays, experimentally testing ~1700 miRNAs without relying on target prediction algorithms. Although not fully representing the secondary structure of the full *HTT* mRNA, our full length 3′-UTR reporter vectors may capture some of local secondary structures produced by 3′-UTR region. As anticipated, many miRNAs that were predicted to target *HTT* did not significantly reduced *HTT* 3-UTR reporter activity in our assays (Table S1). However, 2 candidate miRNAs that were identified from a series of reporter vector activity assays in this study also reduced the endogenous HTT protein levels, supporting that our full length 3′-UTR reporter vector may recapitulate the structure of endogenous *HTT* mRNA 3′-UTR.

Of note, many miRNAs seemed to interact with coding sequence of the Firefly luciferase reporter vector, and we were able to identify such false positives by using empty vector controls and reporter vectors conjugated with different *HTT* haplotypes. Like numerous predicted target sites in *HTT*, coding sequence of reporter vector may have similar sequence as ~7 nucleotide long miRNA seed sequences. Thus, when using reporter vectors for miRNA research, the use of appropriate controls is strongly recommended for correct interpretation of the results. Again, our data did not strongly support miRNAs that were previously reported to interact with *HTT* 3′-UTR and reduce *HTT* expression levels [25, 40]. Many of them were nominated since they were predicted to interact with *HTT* [25, 40] or showed differential expression in HD [22–24]. Because *HTT* 3′-UTR is relatively big, seed sequences of numerous miRNAs are found in the *HTT* 3′-UTR region. For example, the TargetScan prediction algorithm predicted 1694 unique miRNA-target sites for 1074 miRNAs, suggesting that virtually all miRNAs may be able to interact with *HTT* 3′-UTR. Therefore, any miRNA at higher concentrations is likely to suppress *HTT* expression levels in in vitro experiments. Supporting this, 9 miRNAs (miR-214-3p, miR-150-5p, miR-146a-5p, miR-125b-5p, miR-137, miR-148a-3p, miR-128a, miR-196a, miR-27a-3p) previously reported in the literature to reduce *HTT* expression levels [25, 35, 36, 40–42] were not significant by FDR in our primary screening assay except hsa-miR-125b-5p, which was used as a control miRNA in our primary screening assays (Table S1). In addition, the use of reporter vector containing a small region of *HTT* 3′-UTR might have contributed to overestimating the impacts of certain miRNAs. Considering these, our unbiased experiment-based approach using full length *HTT* 3′-UTR produced an objective data that supports various aspects of miRNA research for HD, representing a relevant development and significant advancement.

Currently, it is not clear whether those candidate miRNAs are expressed in brains and interact with 3′-UTR of *HTT* to regulate mRNA levels. Our candidate miRNAs are not described in the brain RNAseq data [22, 23, 39], suggesting low expression levels in brain regions, if any. Nevertheless, one can hypothesize that if used as silencing reagents, our candidate miRNAs may generate strong lowering effects due to the lack of basal activities. Unfortunately, our candidate miRNAs based on full human *HTT* 3′-UTR can't be meaningfully tested in mice due to the lack of models carrying expanded CAG repeat in the endogenous mouse *Htt* locus with full human 3′-UTR sequence. Still, our unbiased data provide important insights into accessible regions in the *HTT* 3′-UTR, and thus follow-up molecular mapping may help develop highly efficient lowering strategies using siRNA or artificial miRNAs. The candidate miRNAs that we discovered can't be used for therapeutics as current format because those miRNAs may target other genes, a phenomenon called multi-targeting [32–34]. It may not be feasible to fully identify all the targets of our candidate miRNAs in brains due to technical difficulty of experiments involving human brains. However, our data clearly demonstrated that endogenous miRNAs that can interact with *HTT* 3′-UTR exist and can be discovered through unbiased approaches. We believe that knowledge of endogenous regulatory mechanisms of *HTT* expression may contribute to developing therapeutic lowering strategies for HD. Thus, our data providing full 3′-UTR sequences of common *HTT* haplotypes, their allelic differences, and demonstration of lowering *HTT* 3′-UTR activities by naturally-occurring miRNAs provide strong support for alternative routes to deal with the root cause of the disease.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. THDCRG. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 1993;72:971–83.

2. Keum JW, Shin A, Gillis T, Mysore JS, Abu Elneel K, Lucente D, et al. The HTT CAG-expansion mutation determines age at death but not disease duration in Huntington disease. Am J Hum Genet 2016;98:287–98.

3. Lee JM, Ramos EM, Lee JH, Gillis T, Mysore JS, Hayden MR, et al. CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion. Neurology 2012;78:690–5.

4. Keiser MS, Kordasiewicz HB, McBride JL. Gene suppression strategies for dominantly inherited neurodegenerative diseases: lessons from Huntington's disease and spinocerebellar ataxia. Hum Mol Genet 2016;25:R53–64.

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

6. 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.

7. Bonini NM, La Spada AR. Silencing polyglutamine degeneration with RNAi. Neuron 2005;48:715–8.

8. Baine FK, Kay C, Ketelaar ME, Collins JA, Semaka A, Doty CN, et al. Huntington disease in the South African population occurs on diverse and ethnically distinct genetic haplotypes. Eur J Hum Genet: Ejhg 2013;21:1120–7.

9. Chao MJ, Gillis T, Atwal RS, Mysore JS, Arjomand J, Harold D, et al. Haplotype-based stratification of Huntington's disease. Eur J Hum Genet: Ejhg 2017;25:1202–9.

10. Warby SC, Montpetit A, Hayden AR, Carroll JB, Butland SL, Visscher H, et al. CAG expansion in the Huntington disease gene is associated with a specific and targetable predisposing haplogroup. Am J Hum Genet 2009;84:351–66.

11. Goldberg YP, Kremer B, Andrew SE, Theilmann J, Graham RK, Squitieri F, et al. Molecular analysis of new mutations for Huntington's disease: intermediate alleles and sex of origin effects. Nat Genet 1993;5:174–9.

12. Lee JM, Kim KH, Shin A, Chao MJ, Abu Elneel K, Gillis T, et al. Sequence-level analysis of the major European Huntington disease haplotype. Am J Hum Genet 2015;97:435–44.

13. Myers RH, MacDonald ME, Koroshetz WJ, Duyao MP, Ambrose CM, Taylor SA, et al. De novo expansion of a (CAG)n repeat in sporadic Huntington's disease. Nat Genet 1993;5:168–73.

14. Semaka A, Collins JA, Hayden MR. Unstable familial transmissions of Huntington disease alleles with 27-35 CAG repeats (intermediate alleles). Am J Med Genet Part B, Neuropsychiatr Genet 2010;153B:314–20.

15. Semaka A, Kay C, Belfroid RD, Bijlsma EK, Losekoot M, van Langen IM, et al. A new mutation for Huntington disease following maternal transmission of an intermediate allele. Eur J Med Genet 2015;58:28–30.

16. Lee JM, Gillis T, Mysore JS, Ramos EM, Myers RH, Hayden MR, et al. Common SNP-based haplotype analysis of the 4p16.3 Huntington disease gene region. Am J Hum Genet 2012;90:434–44.

17. Carroll JB, Warby SC, Southwell AL, Doty CN, Greenlee S, Skotte N, et al. Potent and selective antisense oligonucleotides targeting single-nucleotide polymorphisms in the Huntington disease gene/allele-specific silencing of mutant huntingtin. Mol Ther J Am Soc Gene Ther 2011;19:2178–85.

18. Southwell AL, Kordasiewicz HB, Langbehn D, Skotte NH, Parsons MP, Villanueva EB, et al. Huntingtin suppression restores cognitive function in a mouse model of Huntington's disease. Sci Transl Med. 2018;10. pii: 10/461/eaar3959. https://doi.org/10.1126/scitranslmed.aar3959.

19. Shin JW, Kim KH, Chao MJ, Atwal RS, Gillis T, MacDonald ME, et al. Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9. Hum Mol Genet 2016;25:4566–76.

20. Harper SQ, Staber PD, He X, Eliason SL, Martins IH, Mao Q, et al. RNA interference improves motor and neuropathological abnormalities in a Huntington's disease mouse model. Proc Natl Acad Sci USA 2005;102:5820–5.

21. Boudreau RL, McBride JL, Martins I, Shen S, Xing Y, Carter BJ, et al. Nonallele-specific silencing of mutant and wild-type huntingtin demonstrates therapeutic efficacy in Huntington's disease mice. Mol Ther J Am Soc Gene Ther 2009;17:1053–63.

22. Hoss AG, Kartha VK, Dong X, Latourelle JC, Dumitriu A, Hadzi TC, et al. MicroRNAs located in the Hox gene clusters are implicated in huntington's disease pathogenesis. PLoS Genet 2014;10:e1004188.

23. Marti E, Pantano L, Banez-Coronel M, Llorens F, Minones-Moyano E, Porta S, et al. A myriad of miRNA variants in control and Huntington's disease brain regions detected by massively parallel sequencing. Nucleic acids Res 2010;38:7219–35.

24. Packer AN, Xing Y, Harper SQ, Jones L, Davidson BL. The bifunctional microRNA miR-9/miR-9* regulates REST and CoREST and is downregulated in Huntington's disease. The. J Neurosci 2008;28:14341–6.

25. Sinha M, Ghose J, Bhattarcharyya NP. Micro RNA -214,-150,-146a and-125b target Huntingtin gene. RNA Biol 2011;8:1005–21.

26. Reed ER, Latourelle JC, Bockholt JH, Bregu J, Smock J, Paulsen JS, et al. MicroRNAs in CSF as prodromal biomarkers for Huntington disease in the PREDICT-HD study. Neurology 2018;90:e264–72.

27. Ambrose CM, Duyao MP, Barnes G, Bates GP, Lin CS, Srinidhi J, et al. Structure and expression of the Huntington's disease gene: evidence against simple inactivation due to an expanded CAG repeat. Somat cell Mol Genet 1994;20:27–38.

28. 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.

29. 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.

30. Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. Widespread changes in protein synthesis induced by microRNAs. Nature 2008;455:58–63.

31. Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. Nature 2008;455:64–71.

32. Krutzfeldt J, Poy MN, Stoffel M. Strategies to determine the biological function of microRNAs. Nat Genet 2006;38(Suppl):S14–9.

33. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, et al. Combinatorial microRNA target predictions. Nat Genet 2005;37:495–500.

34. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. Cell 2003;115:787–98.

35. Fu MH, Li CL, Lin HL, Tsai SJ, Lai YY, Chang YF, et al. The potential regulatory mechanisms of miR-196a in Huntington's disease through bioinformatic analyses. PloS ONE 2015;10:e0137637.

36. Kocerha J, Xu Y, Prucha MS, Zhao D, Chan AW. microRNA-128a dysregulation in transgenic Huntington's disease monkeys. Mol brain. 2014;7:46.

37. Sinha M, Mukhopadhyay S, Bhattacharyya NP. Mechanism(s) of alteration of micro RNA expressions in Huntington's disease and their possible contributions to the observed cellular and molecular dysfunctions in the disease. Neuromolecular Med. 2012;14: 221–43.

38. Basak I, Patil KS, Alves G, Larsen JP, Moller SG. microRNAs as neuroregulators, biomarkers and therapeutic agents in neurodegenerative diseases. Cell Mol life Sci. 2016;73:811–27.

39. Hoss AG, Labadorf A, Latourelle JC, Kartha VK, Hadzi TC, Gusella JF, et al. miR-10b-5p expression in Huntington's disease brain relates to age of onset and the extent of striatal involvement. BMC Med Genom. 2015;8:10.

40. Kozlowska E, Krzyzosiak WJ, Koscianska E. Regulation of huntingtin gene expression by miRNA-137, -214, -148a, and their respective isomiRs. Int J Mol Sci. 2013;14:16999–7016.

41. Ban JJ, Chung JY, Lee M, Im W, Kim M. MicroRNA-27a reduces mutant hutingtin aggregation in an in vitro model of Huntington's disease. Biochem Biophys Res Commun. 2017;488:316–21.

42. Cheng PH, Li CL, Chang YF, Tsai SJ, Lai YY, Chan AW, et al. miR-196a ameliorates phenotypes of Huntington disease in cell, transgenic mouse, and induced pluripotent stem cell models. Am J Hum Genet 2013;93:306–12.

43. Jovicic A, Zaldivar Jolissaint JF, Moser R, Silva Santos Mde F, Luthi-Carter R. MicroRNA-22 (miR-22) overexpression is neuroprotective via general anti-apoptotic effects and may also target specific Huntington's disease-related mechanisms. PloS ONE 2013;8:e54222.

44. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. Nature 2005;435:834–8.

45. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative genomic profiling of human prostate cancer. Cancer cell 2010;18:11–22.

46. Johnson R, Buckley NJ. Gene dysregulation in Huntington's disease: REST, microRNAs and beyond. Neuromolecular Med 2009;11:183–99.

47. Gurtan AM, Sharp PA. The role of miRNAs in regulating gene expression networks. J Mol Biol. 2013;425:3582–600.

48. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell 2009;136:215–33.

49. Grun D, Wang YL, Langenberger D, Gunsalus KC, Rajewsky N. microRNA target predictions across seven Drosophila species and comparison to mammalian targets. PLoS Comput Biol. 2005; 1:e13.

50. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005;120:15–20.