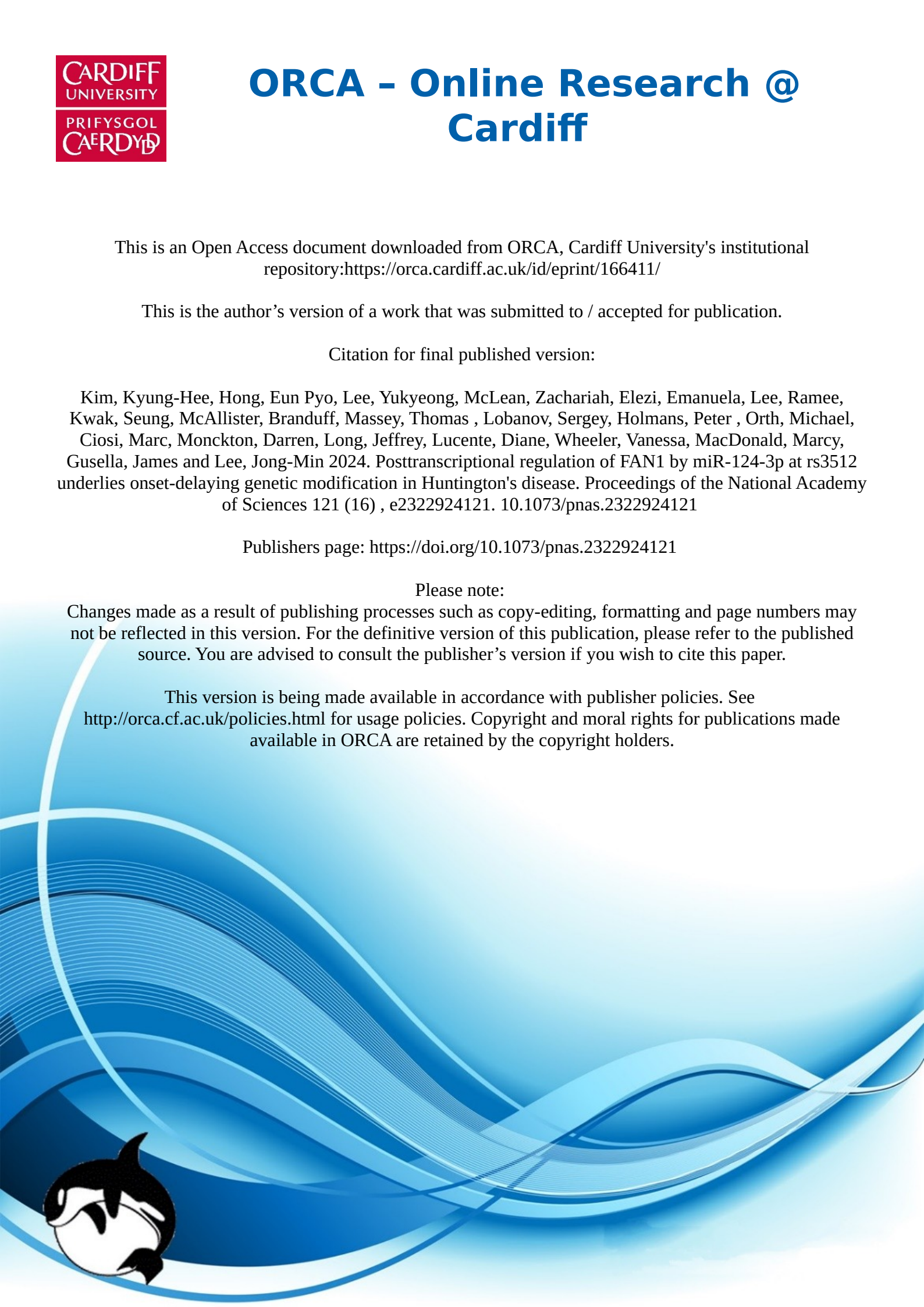


Main Manuscript for

Post-transcriptional regulation of *FAN1* by miR-124-3p at rs3512 underlies onset-delaying genetic modification in Huntington's disease

Kyung-Hee Kim[1,2], Eun Pyo Hong[1,2], Yukyeong Lee[1,2], Zachariah L. McLean[1,2,8], Emanuela Elezi[1], Ramee Lee[3], Seung Kwak[3], Branduff McAllister[1,4], Thomas H. Massey[4], Sergey Lobanov[4], Peter Holmans[4], Michael Orth[5], Marc Ciosi[6], Darren G. Monckton[6], Jeffrey D. Long[7], Diane Lucente[1], Vanessa C. Wheeler[1,2], Marcy E. MacDonald[1,2,8], James F. Gusella[1,8,9], and Jong-Min Lee[1,2,8,*]

[1] 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] Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff CF24 4HQ, UK

[5] University Hospital of Old Age Psychiatry and Psychotherapy, Bern University, Bern, Switzerland

[6] School of Molecular Biosciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK

[7] Department of Psychiatry, Carver College of Medicine and Department of Biostatistics, College of Public Health, University of Iowa, Iowa City, Iowa 52242, USA

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

[9] Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02115, USA

* Jong-Min Lee

Email: jlee51@mgh.harvard.edu

Author Contributions: K.H.K and J.M.L designed the study. K.H.K, E.P.H, Y.L, Z.K.N, and E.E performed experiments. K.H.K and J.M.L analyzed data. R.L, S.K, B.M, T.H.M, S.L, P.H, M.O, M.C, D.G.M, J.D.L, D.L, V.C.W, M.E.M, and J.F.G generated and provided resources for this

study. K.H.K and J.M.L wrote the manuscript with the help of R.L, S.K, B.M, T.H.M, S.L, P.H, M.O, M.C, D.G.M., J.D.L, D.L, V.C.W, M.E.M, and J.F.G.

Competing Interest Statement: J.M.L. consults for Life Edit Therapeutics and serves on the scientific advisory board of GenEdit Inc. J.F.G. was a Founder and Scientific Advisory Board member with a financial interest in Triplet Therapeutics Inc. His NIH-funded project is using genetic and genomic approaches to uncover other genes that significantly influence when diagnosable symptoms emerge and how rapidly they worsen in Huntington's disease. The company was developing new therapeutic approaches to address triplet repeat disorders such as Huntington's disease, myotonic dystrophy, and spinocerebellar ataxias. Dr. Gusella's interests were reviewed and are managed by Massachusetts General Hospital and Mass General Brigham in accordance with their conflict of interest policies. Dr. Gusella also consults for Transine Therapeutics Ltd and has been a consultant for Wave Life Sciences USA Inc., Biogen Inc., and Pfizer Inc.. V.C.W. was a Scientific Advisory Board member of Triplet Therapeutics, Inc., a company developing new therapeutic approaches to address triplet repeat disorders such Huntington's disease and myotonic dystrophy. Her financial interests in Triplet Therapeutics were reviewed and are managed by Massachusetts General Hospital and Mass General Brigham in accordance with their conflict of interest policies. She is a scientific advisory board member of LoQus23 Therapeutics and has provided paid consulting services to Alnylam, Acadia Pharmaceuticals Inc., Alnylam Inc., Biogen Inc. and Passage Bio. S.L. is currently employee of Illumina, Inc., a public company that develops and markets systems for genetic analysis. Within the last five years Professor Monckton has been a scientific consultant and/or received an honoraria/grants from AMO Pharma, Dyne, F. Hoffman-La Roche, LoQus23, MOMA Therapeutics, Novartis, Ono Pharmaceuticals, Pfizer Pharmaceuticals, Rgenta Therpeutics, Sanofi, Sarepta Therapeutics Inc, Script Biosciences, Triplet Therapeutics, and Vertex Pharmaceuticals. Professor Monckton also had research contracts with AMO Pharma and Vertex Pharmaceuticals. T.H.M. is an associate member of the scientific advisory board of LoQus23 Therapeutics Ltd.

Classification: Biological Sciences; Genetics

Keywords: Huntington's disease; genetic modifier; *FAN1*; repeat instability; rs3512; miR-124-3p

This PDF file includes:

Main Text
Figures 1 to 4

Abstract

Many Mendelian disorders, such as Huntington's disease (HD) and spinocerebellar ataxias (SCAs), arise from expansions of CAG trinucleotide repeats. Despite the clear genetic causes, additional genetic factors may influence the rate of those monogenic disorders. Notably, genome-wide association studies discovered somewhat expected modifiers, particularly mismatch repair genes involved in the CAG repeat instability, impacting age-at-onset of HD. Strikingly, *FAN1*, previously unrelated to repeat instability, produced the strongest HD modification signals. Diverse *FAN1* haplotypes independently modify HD, with rare genetic variants diminishing DNA binding or nuclease activity of the FAN1 protein, hastening HD onset. However, the mechanism behind the frequent and the most significant onset-delaying *FAN1* haplotype lacking missense variations has remained elusive. Here, we illustrated that a microRNA (miRNA) acting on 3'-UTR SNP rs3512, rather than transcriptional regulation, is responsible for the significant *FAN1* eQTL (expression quantitative trait loci) signal and allelic imbalance in *FAN1* mRNA, accounting for the most significant and frequent onset-delaying modifier haplotype in HD. Specifically, miR-124-3p selectively targets the reference allele at rs3512, diminishing the stability of *FAN1* mRNA harboring that allele and consequently reducing its levels. Subsequent validation analyses, including the use of antagomir and 3'-UTR reporter vectors with swapped alleles, confirmed the specificity of miR-124-3p at rs3512. Together, these findings indicate that the alternative allele at rs3512 renders the *FAN1* mRNA less susceptible to miR-124-3p-mediated post-transcriptional regulation, resulting in increased FAN1 levels and a subsequent delay in HD onset by mitigating CAG repeat instability.

Significance Statement

Expansions of CAG trinucleotide repeats cause many Mendelian disorders, such as Huntington's disease (HD). While mismatch repair genes involved in repeat instability were expected to modify HD, *FAN1*, not previously linked to instability, emerged as the strongest modifier of HD. A frequent *FAN1* haplotype, lacking missense variations, represents the most significant onset-delaying modifier in HD. Surprisingly, frequent *FAN1*-mediated delay was attributed to a microRNA (miR-124-3p) acting on a 3'-UTR SNP (rs3512) in *FAN1* rather than transcriptional regulation. The alternative allele at rs3512 reduces sensitivity to miR-124-3p, leading to elevated FAN1 levels and ultimately delaying HD onset by reducing CAG repeat instability. Our data revealed relevant target mechanisms supported by patient genetics and thus may offer new opportunities for disease-modifying therapeutics.

Main Text

Introduction

Expansions of short tandem repeat DNA cause many Mendelian disorders (1, 2). However, those single gene disorders are also significantly modified by other genetic factors. For example, the expanded CAG trinucleotide repeat in the huntingtin gene (*HTT)* results in the neurological abnormalities of Huntington's disease (HD; MIM #143100), with nearly complete penetrance with inherited repeats longer than 39 CAGs (3, 4). However, there is a quantitative nature to these effects as the age at disease manifestation is inversely correlated with the size of the inherited *HTT* CAG repeat (5-9). Genetic investigation of this genotype-phenotype relationship has revealed that age-at-onset is 1) largely determined by the size of uninterrupted CAG repeat, rather than the polyglutamine segment of huntingtin that it encodes, and 2) modified by genetic factors elsewhere in the genome (10-13). Importantly, many HD onset modifier loci encode DNA repair genes involved in somatic CAG repeat instability (10-16), implicating this process in the timing of HD manifestations.

Most of the HD modifier loci are involved in DNA mismatch repair (MMR), but the most significant onset modification signals in the genome-wide association studies (GWAS) occur at *FAN1,* an interstrand DNA-crosslink repair (ICL) gene (10, 12, 17-19) that has also been associated with altered *HTT* CAG repeat instability (13, 16, 20-25). Interestingly, the association signals at *FAN1* involve multiple independent modifier effects, including infrequent onset-hastening (i.e., 15AM1 and 15AM3) and common onset-delaying haplotypes (i.e., 15AM2) (12). Reduced DNA binding/nuclease activity of FAN1 due to missense variants rs151322829 and rs150393409 account for the onset-hastening effects of 15AM1 and 15AM3 (21, 22). In contrast, no *FAN1* coding variant tags the frequent onset-delaying effect of the 15AM2 haplotype (12). Instead, increased *FAN1* expression levels in the cortex appear to delay age-at-onset in HD (12, 21) by stabilizing the *HTT* CAG repeat (20, 23). The onset-delaying *FAN1* haplotype is quite common; the allele frequency of the top tagging SNP indicates that approximately 50% of HD patients carry at least one 15AM2 modifier haplotype (12). Still, the underlying mechanism of increased *FAN1* expression level from the 15AM2 haplotype has remained unclear, hampering development of *FAN1*-based disease-delaying treatments. In view of the lack of any disease-modifying treatment for HD, understanding the mechanism by which 15AM2 *FAN1* delay HD may point to novel therapeutic strategies for this devastating disorder.

Here, we conducted thorough genetic analyses and subsequent molecular validation experiments to elucidate the genetic basis underlying the pronounced delay of HD onset attributed to *FAN1*. Our investigations revealed that a specific genetic variant, rs3512, induces a

sequence alteration in the binding site for miR-124-3p, resulting in a selective reduction of *FAN1* mRNA. The consequential elevation of FAN1 levels in patient-derived cells carrying the rs3512 alternative allele led to the stabilization of the *HTT* CAG repeat, providing insights into underlying mechanisms of the frequent and most significant onset delaying effect in HD. These findings represent the first elucidation of a disease-modifying mechanism, wherein a miRNA pathway interacts with a gene crucial for maintaining CAG repeat stability, informing disease-modifying therapeutic strategies for HD and potentially other repeat expansion disorders.

Results

The most significant onset-delaying HD modifier represents *FAN1* eQTL that does not involve transcriptional regulation.

To identify the causal variant responsible for the most significant onset-delaying modifier in HD, we focused on 112 tagging SNPs that are in linkage disequilibrium (LD) (S. Fig. 1; open blue circles) with the lead SNP (rs35811129; a filled red circle in S. Fig. 1A) that marked the most significant onset-delaying *FAN1* haplotype, namely 15AM2 *FAN1* haplotype (12). Subsequent fine-mapping analysis (26) of 15AM2-tagging SNPs narrowed the list to a credible set of 17 SNPs representing potential causal variants (S. Table 1). However, it failed to pinpoint a single causal variant for 15AM2 modification effect due to strong LD (linkage disequilibrium) (S. Fig. 1B). Since none of the 15AM2-tagging SNPs or credible set SNPs alters protein sequence (S. Table 1), and co-localization analysis suggested overlap of 15AM2 modification signals and *FAN1* eQTLs (21), we determined which 15AM2-tagging SNPs could explain the *FAN1* expression levels. As shown in Fig. 1A, genome-wide eQTL analysis for *FAN1* using the CommonMind Consortium (CMC) RNAseq data revealed significant *cis* eQTL signals at the chromosome 15 without significant *trans* eQTL or long-range enhancer signals (Fig. 1B). As expected, *FAN1* eQTL signals (Fig. 1B) and HD modification signals for 15AM2 (Fig. 1C) were strongly correlated (Fig. 1C; Pearson's correlation p-value, 5.1E-28), supporting a role for altered *FAN1* expression levels in the modification of HD.

Since transcription plays a pivotal role in gene expression, we further investigated whether HD modification due to *FAN1* eQTL was mediated by transcriptional regulation. Specifically, we examined the relationship between the ENCODE regulation marks and the HD modification signals for 15AM2 *FAN1*. As summarized in S. Fig. 2, 15AM2-tagging SNPs are located across *FAN1*, but they did not show significant correspondence with the sites of major histone marks (H3K4Me1, H3K4Me3, H3K27Ac) (secondary Y-axis on the top panel),

transcription factor ChIP-sequence signals (primary Y-axis on the middle panel), or DNase I hypersensitivity regions (secondary Y-axis in the middle panel). Indeed, most of the SNPs in the credible set are located downstream of the *FAN1* coding region, around 3'-UTR (untranslated region) (Fig. 1C; S. Fig. 2). Reflecting this lack of correspondence, correlation analysis did not reveal significant relationships between onset modification signals and the ENCODE regulation marks for the 15AM2 modifier haplotype (S. Fig. 3). Similarly, *FAN1* eQTL signals were not correlated with the ENCODE regulation marks (data not shown). Furthermore, the basal promoter activity of 15AM2 *FAN1* was not significantly different from that of a non-15AM2 promoter (S. Fig. 4; p-value, 0.663), arguing against a role for transcriptional regulation in the increased expression levels in 15AM2 *FAN1*.

Significant onset modification, eQTL, and allelic imbalance for the 3'-UTR SNP rs3512.

Since the increased *FAN1* levels due to 15AM2 couldn't be explained by transcriptional regulation, and the credible set SNPs were centered around the 3'-UTR region of *FAN1*, we evaluated the alternative hypothesis that post-transcriptional regulation was responsible for the 15AM2 modifier effect. To evaluate this possibility, we also determined the allelic expression of the exonic variants through allele-specific expression analysis (ASE). Due to significant alignment bias in ASE analysis (27), we compared the observed ratio of alternative and reference alleles to the expected ratio to determine imbalanced expression of *FAN1* mRNA. Among 112 tagging SNPs, three SNPs are located in the exons of *FAN1*; only a 3'-UTR SNP, rs3512, which was one of the credible set in the fine-mapping analysis, showed significant signals in ASE analysis, onset modification, and eQTL analyses (Fig. 2A). The alternative allele of rs3512, which is associated with delayed onset, is significantly overrepresented in both CMC (prefrontal cortex; Fig. 2B) and GTEx RNAseq data (S. Fig. 5), indicating that individuals heterozygous at rs3512 express the alternative allele at a higher level than the reference allele. Because 1) the sequence of 15AM2 haplotype supported increased *FAN1* expression levels without altering transcriptional regulation, 2) there is no significant splicingQTL for *FAN1* in the relevant brain regions (GTEx and BrainMeta) (28), 3) the 3'-UTR SNP rs3512 tagging the 15AM2 haplotype shows strong HD modification, eQTL, and ASE signals, 4) public RNAseq repositories show a higher than expected relative level of the rs3512 alternative allele in the cortex, and 5) miRNAs have been shown to regulate the expression levels of target mRNAs by altering their stabilities (29, 30), we, therefore, hypothesized that allele-specific targeting of rs3512 by miRNA alters *FAN1* mRNA levels through post-transcriptional regulation.

miR-124-3p preferentially targets the reference allele of rs3512 in the *FAN1* mRNA.

To determine whether *the FAN1* mRNA region involving rs3512 is targeted by a miRNA, we queried miRNA target prediction algorithms. The TargetScan program (31) revealed that miRNA hsa-miR-124-3p (miR-124-3p hereafter), which is highly expressed in brain (S. Fig. 6), was predicted to target the reference allele at rs3512 (non-15AM2) (Fig. 2C; red). Due to the mismatch in the seed sequence, this miRNA was anticipated to target 15AM2-derived *FAN1* mRNA less efficiently (Fig. 2C; blue). Consistent with these predictions, miR-124-3p significantly decreased endogenous FAN1 protein levels in HEK293T cells, which carry two reference alleles at rs3512 (Fig. 2D; p-value, 4.1E-3). Subsequent MiSeq analysis of an HD patient-derived lymphoblastoid cell line (LCL) which is heterozygous at rs3512 showed that the expected slight underrepresentation of the alternative allele due to alignment bias ('Scramble') was reversed by miR-124-3p (Fig. 2E; p-value, 2.9E-5). Furthermore, 3'-UTR luciferase reporter vector assays (S. Fig. 7) showed that miR-124-3p resulted in a modest reduction of the reporter activity of 15AM2 (Fig. 2F; blue), while the same miRNA significantly decreased the luciferase activity of the non-15AM2 reporter in a dose-dependent manner (Fig. 2F; red).

In addition, the effects of miR-124-3p on the non-15AM2 3'-UTR were abolished by a specific inhibitor of miR-124-3p. For example, miR-124-3p resulted in modest and significant reduction of 3'-UTR reporter activities for 15AM2 (blue) and non-15AM2 (red), respectively; the antagomir for miR-124-3p blocked the effects of miR-124-3p completely (Fig. 3A). Furthermore, the antagomir for miR-124-3p reversed 1) the overrepresentation of the alternative allele at rs3512 induced by miR-124-3p in HD neural progenitor cells (NPCs) carrying 15AM2/non-15AM2 haplotypes (Fig. 3B) and 2) the reduction in FAN1 protein levels due to miR-124-3p (Fig. 3C). We then determined whether the impact of miR-124-3p was specifically due to rs3512 by testing 3'-UTR reporter vectors with swapped alleles at rs3512 (Fig. 3D). When the reference allele of rs3512 in the non-15AM2 3'-UTR reporter was replaced with the alternative allele (non-15AM2-C), miR-124-3p no longer decreased the reporter activity (Fig. 3E). In contrast, miR-124-3p significantly decreased the reporter activity of 15AM2-G carrying the reference allele at rs3512 (Fig. 3E), indicating that the allele-specific effect of miR-124-3p is mediated specifically by the rs3512 site.

miR-124-3p reduces the stability of *FAN1* mRNA resulting in the increase of CAG repeat expansion.

HTT CAG repeat instability represents the primary HD onset modifier (12). In addition, knock-out and over-expression of *FAN1* respectively increase and decrease the rate of somatic CAG repeat expansion (16, 20, 21). Therefore, we tested whether miR-124-3p could increase CAG repeat expansion. Since repeat expansion could be detected in cultured cells when the CAG repeats were very long (20), we analyzed a patient-derived NPC line carrying 130 CAGs to increase the

chance of generating time-dependent CAG repeat expansion *in vitro*. Unfortunately, these cells carry homozygous alternative alleles of rs3512, implying that the effects of miR-124-3p in these cells would be modest. Owing to the long CAG repeat in these cells, a time-dependent increase in expansion index was detected in these cells (Fig. 4A; black; p-value, 1.8E-9). Although not very dramatic, potentially due to the *FAN1* haplotypes in the cells that we used, miR-124-3p increased the levels of repeat expansion further (Fig. 4A; gold; p-value, 4.0E-3). Since miR-124-3p significantly decreased the total *FAN1* mRNA levels (Fig. 4B; p-value, 2.3E-3) and increased the rs3512 alternative/reference ratio in the *FAN1* mRNA in the presence of actinomycin D (Fig. 4C; p-value, 1.4E-7), these data suggest that different alleles of rs3512 confer quantitatively different stability on *FAN1* mRNA, altering its levels through miR-124-3p-mediated post-transcriptional regulation.

Underlying mechanism of reduced significance for rs3512 in the GWAS.

It is noteworthy that 1) our fine-mapping analysis did not specify a single causal variant for the 15AM2 modifier effect and 2) the putative causal variant for 15AM2 haplotype (i.e., rs3512) did not generate the highest association signal in our GWAS. Rather, a linked SNP, rs35811129, produced the strongest age-at-onset modification signal. This was not because multiple causal variants are responsible for the 15AM2 effect but because the strong infrequent onset-hastening 15AM3 haplotype (rs151322829) is more associated with rs3512 (S. Fig. 8). For example, LD between rs3512-rs151322829 (R-square, 0.019) was higher than that of rs35811129-rs151322829 (R-square, 0.0032). Therefore, age-at-onset analysis for rs3512 was influenced by the presence of onset-hastening rs151322829 (i.e., causal variant of the 15AM3 onset-hastening modifier effect) in a few individuals, thereby diminishing the overall association significance of rs3512. In support, rs3512 produced the strongest onset modification signal when 1) the statistical model excluded a small number of HD subjects carrying 15AM3 (S. Fig. 8C) and 2) an association model was corrected for the effect of 15AM3 by including rs151322829 as a covariate. These data illustrate that the existence of multiple effect haplotypes confounded the single SNP association analysis and our fine-mapping analysis to identify the causal variant for 15AM2. We anticipate that this phenomenon may occur in GWAS of other diseases, making the delineation of the causal variants very difficult, particularly when local LD is strong.

Discussion

Our genetic investigation focusing on the strongest HD onset modification signals revealed that miR-124-3p targets the *FAN1* mRNA harboring rs3512 alternative allele less efficiently due to a mismatch between miRNA and target. Since a significant proportion of miRNA-target interactions (~30%) contains a single mismatch in the seed region (32) and our data showed that miR-124-3p still targets the 15AM2 weakly, different alleles of rs3512 may generate quantitatively different impacts on the stability of *FAN1* mRNA, accounting for the frequent onset-delaying effect of 15AM2 *FAN1* haplotype.

Often, genetic associations with human traits are discovered for SNPs located in non-coding regions that involve many linked variants in strong LD, making identification of a single causal variation quite challenging. Recently, massively parallel reporter assays (MPRAs) were applied to identify causal variants for eQTLs, revealing that regulatory variants are specifically enriched for activating chromatin structures and allelic transcription factor binding (33). In contrast, rs3512, which tags the 15AM2 modifier haplotype and was one of the credible set in the fine-mapping analysis, does not produce allelic effects on *FAN1* expression levels by altering transcriptional regulation. Rather, rs3512 generates an eQTL signal by changing the target seed sequence of miR-124-3p, representing the first demonstration of the modulation of repeat instability and modification of HD by a miRNA pathway. Relatively increased targeting at the rs3512-reference allele by miR-124-3p destabilizes the *FAN1* mRNA, leading to a relatively higher level of *FAN1* expression in HD individuals carrying the rs3512 alternative allele compared to subjects carrying the alternative allele. Reduced FAN1 function is correlated with increased CAG expansion (16, 20-22), so our data argue that the quantitatively differential targeting efficiencies of miR-124-3p at rs3512 influence the *FAN1* levels, and therefore, alter the levels of somatic repeat expansion of the *HTT* CAG repeat. The relative increase in *FAN1* expression in individuals carrying the rs3512-alternative allele may shift the competitive balance between the activities of FAN1 and MMR to favor the FAN1 pathway (Fig. 4D and 4E).

Defective DNA repair due to mutations in *FAN1* can result in chronic kidney failure and hereditary colorectal cancer (34, 35). The rs3512 is not associated with these diseases but has been implicated in age-at-onset of spinocerebellar ataxias (36) in addition to HD (this study). Thus, the physiological variance in *FAN1* expression levels due to rs3512 does not cause disease but can potentially influence the rate of pathogenesis in diseases where DNA repair plays an important role. Interestingly, *FAN1* may not modify all repeat expansion disorders, as age-associated penetrance in X-linked dystonia-parkinsonism caused by an expanded CCCTCT repeat is modified by *MSH3* and *PMS2*, but not by *FAN1* (37). These data suggest a possibility that that certain DNA repair genes such as *MSH3* and *PMS2* are broadly involved in repeat

instability, while *FAN1* may be more specific to the maintenance of the stability of certain tandem repeats (37, 38).

A role for the brain-specific miR-124-3p in neurogenesis and differentiation has been described in the literature (39-42). For instance, miR-124-3p improved infarct volume, pro-inflammatory indications, and neurogenesis in stroke models (43-46). Also, it inhibited the □-amyloid abnormality and improved cognitive outcomes in traumatic brain injury, which is a risk factor for long-term neurodegenerative disease (47). miR-124-3p was found to be down-regulated in HD brains (48-52), prompting a pre-clinical study to test its efficacy in the R6/2 mouse model, but it did not produce significant behavioral improvement (53). This suggests that reduced levels of this neuron-specific miRNA in HD brains reflects neuronal loss and that the R6/2 phenotype may not require further CAG repeat expansion. Instead, we hypothesize that blocking of the specific interaction between miR-124-3p and the *FAN1* 3'-UTR delays onset in HD by increasing *FAN1* levels since CAG repeat expansion was significantly increased by miR-124-3p in patient-derived cells. Unfortunately, the test of miR-124-3p *in vivo* was not feasible because the sequence identity between the human *FAN1* 3'-UTR and the mouse *Fan1* 3'-UTR is low. Furthermore, the magnitude of repeat instability increase by miR-124-3p in patient-derived cells was modest (Figure 5A), which was somewhat anticipated given the *FAN1* haplotypes of test cells. Therefore, considering the potential of this pathway, the development of model systems that carry human 3'-UTR sequence, express abundant miR-124-3p and show CAG repeat expansion will significantly facilitate the identification of promising treatments.

In summary, identification of relevant targets for disease-modifying therapeutics is increasingly important in HD, considering the lack of effective treatments and recent failures of highly anticipated huntingtin-lowering clinical trials (https://www.hda.org.uk/media/4418/novartis-vibrant-hd-community-letter-final-pdf.pdf) (54-56). Since the success rate of intervention strategies rooted in human genetics is significantly increased (53, 57), delineation of HD modifiers through investigation of individuals with the disease can provide alternative targets. Our data demonstrate that differential targeting efficiencies of miR-124-3p on rs3512 account for the frequent onset-delaying effects of *FAN1* in HD individuals. Examples of interaction between miRNAs and causal variants have been reported in cancer (58) and Friedreich ataxia (59). Our data represent the first delineation of miRNA-mediated regulation of a genetic modifier that resulted in the delay of the disease onset in HD patients. Elucidating the mechanism of genetic modification, as we have done here for the miR-124-3p-rs3512 interaction, which is responsible for the frequent onset-delaying modifier effect at *FAN1*, is thus an important step toward the development of rational treatments for this devastating disorder.

Materials and Methods

SNP association analysis for chromosome 15 region

SNP association analysis for residual age-at-onset phenotype using mixed effect models are described elsewhere (9009 samples) (12, 60). Subject consents and the overall study approval for the GWAS were described elsewhere (12). For subsequent single SNP association analysis focusing on *FAN1*, we performed linear regression analysis using a fixed effects model with the same set of covariates as described previously (12). SNPs in the chr15:31000000-31400000 region (GRCh37/hg19) were analyzed. Genomic coordinate was based on GRCh37/hg19.

Identification of 15AM2-tagging SNPs based on the LD with the top SNP tagging 15AM2 *FAN1*

In order to identify a set of SNPs that tag the effects of frequent onset-delaying modifier effect of *FAN1*, we calculated the LD of SNPs (in the chr15:31000000-31400000 region) with the lead SNP (i.e., rs35811129) using the PLINK program (https://zzz.bwh.harvard.edu/plink/ld.shtml). Subsequently, 112 SNPs with LD R-square values higher than 0.5 were considered as 15AM2-tagging SNPs.

Fine-mapping analysis of the 15AM2-tagging SNPs

To gain insights into causal variation of 15AM2 onset-delaying modification effect, we performed fine-mapping analysis focusing on 112 tagging SNPs using an R package called susieR (https://cran.r-project.org/web/packages/susieR/index.html). Briefly, genotype data and association analysis results for the region (19) were used for the univariate regression function in the susieR package with 'L=1' option since we focused on single effect tagged by 112 SNPs. This analysis yielded a credible set of 17 SNPs with coverage of 0.9514, meaning the proportion of credible set that contain a true causal SNP.

Genome-wide eQTL analysis for *FAN1* using CMC data

To judge the levels of association of SNPs with the expression levels of *FAN1* in the cortex, genome-wide eQTL analysis was performed using CMC data set (https://www.synapse.org//#!Synapse:syn2759792/wiki/69613). Briefly, we identified 451 study participants with European ancestry based on the genotype-based principal components analysis, and subsequently analyzed for the association analysis using the *FAN1* expression levels as the primary phenotype. We included the study group, sex, diagnosis, age-at-death, brain weight, postmortem interval, and four principal components based on RNAseq expression data as covariates to correct potential confounding effects in the association analysis model using

the PLINK program. Due to the small sample size, we focused on SNPs with minor allele frequencies greater than 1%. The levels of correlation between *FAN1* eQTL signals and HD modification signals were determined by Pearson's correlation analysis focusing on the 15AM2-tagging SNPs. Among 112 tagging SNPs, 10 indel sites were not analyzed in the CMC data, and therefore, the correlation analysis was based on 102 SNPs.

Comparison of HD onset modification signals to transcriptional regulation marks

To judge whether onset modification signals at chromosome 15 could be explained by the transcriptional regulation of *FAN1*, we analyzed the ENCODE regulation tracks (https://www.encodeproject.org/) available at the USCS Genome Browser (https://genome.ucsc.edu/index.html). We focused on H3K4Me1 mark (often found near the regulatory elements), H3K4Me3 mark (often found near the promoter), H3K27Ac mark (often found near active regulatory elements), DNaseI hypersensitivity clusters, and transcription factor ChIP-seq cluster. For each mark, the mean values of 7 different cell lines were primarily used to compare to HD onset modification signals from our GWAS. Focusing on the 112 tagging SNPs including 17 credible set SNPs, the levels of correlation between the onset modification signals and ENCODE regulation tracks were determined by Pearson's correlation analysis.
ation signals from our GWAS. Focusing on the 112 tagging SNPs including 17 credible set SNPs, the levels of correlation between the onset modification signals and ENCODE regulation tracks were determined by Pearson's correlation analysis.

Cell culture

HEK293T (https://www.atcc.org/products/crl-3216) and *FAN1* knockout (KO) HEK293T cells (21) were cultured in DMEM (Invitrogen) supplemented with 10% fetal bovine serum (Hyclone) and Penicillin-Streptomycin (Invitrogen) at 37°C and 5% CO_2. Two independent HD patient-derived lymphoblastoid cell lines (LCL) carrying 15AM2/non-15AM2 *FAN1* haplotypes (43/18 and 42/15 CAGs) were cultured in RPMI (Sigma-Aldrich) supplemented with 5% fetal bovine serum and Penicillin-Streptomycin at 37°C and 5% CO_2. HD patient-derived induced pluripotent cells (iPSC; 42/15 CAGs) carrying 15AM2/non-15AM2 haplotypes were derived from a LCL carrying 42/15 CAGs by the Harvard Stem cell Institute iPS core facility (Cambridge, MA). An iPSC carrying 126/22 CAGs (NDS00143) was obtained from the NINDS Human Cell and Data repository (https://stemcells.nindsgenetics.org/). iPSC lines were cultured in mTeSR plus kit (Stemcell technologies) at 37°C and 5% CO_2. To develop the neural progenitor cells (NPCs), embryoid body (EB) were generated from dissociated iPSCs and cultured in STEMdiff SMADi neural induction kit (Stemcell technologies). Then, cells were transferred to Matrigel coated plates (Corning) at day 5. Neural rosettes were selected and replated at day 12 and cells were ready to

be passaged at day 17. NPCs were cultured in STEMdiff Neural Progenitor Medium (Stemcell technologies) at 37°C and 5% CO_2.

FAN1 promoter activity assay

To determine the levels of basal promoter activity of 15AM2 *FAN1*, we generated firefly luciferase vectors containing promoter regions of 15AM2 and non-15AM2 *FAN1*. Since H3K27Ac signals are often found near the active regulatory elements, we PCR amplified a 1.8 Kb genomic region that show high levels of H3K27Ac marks (i.e., chr15:31194984-31196866) and cloned into the pGL4.20 firefly luciferase plasmid. Briefly, genomic DNA was extracted from an HD patient-derived LCL line (15AM2/non-15AM2, 42/15 CAGs) by DNeasy Blood & Tissue kit (Qiagen) and amplified by PrimeStar HS DNA polymerase (Takara) using specific primers (5'-CGGTACCTTACCTCCTTCAGGCCAATG-3' and 5'-GGATATCGAGTATTAGAAAAACTGGATGTTCTGAGCAATAAAACAC-3'). The PCR products were digested by NheI and EcoRV; resulting PCR products were inserted into the pGL4.20 vector (Promega). The sequence of amplified promoter region in plasmids were confirmed by Sanger sequencing at the Massachusetts General Hospital Genomics core (Boston, MA). To determine the promoter activity, the 96-well plated *FAN1* KO HEK293T cells (21) were co-transfected with empty vector (i.e., pGL4.20), 15AM2 promoter vector or non-15AM2 promoter vector by Lipofectamine 2000 for 24 hrs. To normalize the transfection efficiency, control Renilla firefly vector (pGL4.74, Promega) was co-transfected. The luciferase activity was determined by Dual-Glo luciferase reagent (Promega) as described previously (21). Briefly, firefly luciferase reagent was added into each well and incubated for 15 minutes. Luminescence signal of each well was measured by MicroLumatPlus (Berthold). Thirty minutes later, we added stop/Renilla reagent into each well, incubated for 15 minutes, and read the luminescence signal. The firefly promoter activity was divided by the corresponding Renilla luciferase signal to obtain normalized firefly luciferase activity.

Determination of imbalanced expression of *FAN1* mRNA

To determine whether two alleles of a given SNP site are equally expressed in *FAN1* mRNA or not, we performed allele-specific analysis (ASE) using CMC and GTEx data sets. For a given test SNP site, we identified samples with heterozygous genotype; for example, 98 and 83 samples are heterozygous at rs3512 among the CMC and GTEx (cortex) participants, respectively. Subsequently, we calculated the percentage of alternative allele for each subject for a given site. Due to systemic alignment bias in RNAseq allele count data (27), direct comparison of allele counts at a given site in a given sample would generate incorrect results. To correct alignment bias in our ASE analysis, we compared observed percentage of alternative allele to that of

expected alternative allele percentage. We empirically calculated the expected alternative allele percentage values for all allele combinations by analyzing SNP sites with read count greater than 10. The expected alternative allele percentage values were used as the reference to determine the level of allelic imbalance of SNPs using t-test. For rs3512, we also evaluated the proportion of alternative allele in the alignment bias-corrected GTEx data (i.e., WASP correction) (61).

In addition, we experimentally determined the imbalanced expression of *FAN1* in a patient-derived cells. A patient-derived LCL line was transfected with miR-124-3p (100 nM, 72 hours) using nucleofector kit V (VCA-1003, Lonza) to determine the effects of the candidate miRNA on allelic expression. To determine the effects of the inhibitor of miR-124-3p on allelic expression, an HD patient-derived NPC line was transfected with miR-124-3p (100 nM) and its inhibitor (100 nM) using human stem cell nucleofector kit 1 (VPH-5012, Lonza, 48 hr). Also, the stability of *FAN1* mRNA carrying the alternative allele or the reference allele was determined in a patient-derived LCL line treated with Actinomycin D (SBR00013, Sigma-Aldrich; 10 □g/ml of for 0, 3, 6 and 9 hours). Total RNA samples were isolated (RNeasy Plus mini kit, Qiagen) from patient-derived cells, and cDNA was synthesized using Superscript III fist strand kit (Invitrogen). The cDNA samples were amplified by Ex Taq DNA polymerase (Takara) using specific primers (5'-GAAATGGATAAATGGGGCTTT-3' and 5'-GGCTCACACGTAACCAAACA-3'). The PCR products were purified by QIAquick PCR purification kit (Qiagen) and subjected to the MiSeq analysis at the Massachusetts General Hospital CCIB DNA core (Cambridge, MA). The number of sequence reads for the rs3512 alternative alleles was divided by total read number to calculate the proportion of the alternative allele. Alternatively, the ratio of alternative / reference allele was calculated.
e was calculated.

Prediction of miRNAs that target the rs3512

Candidate miRNA was predicted using the reference sequence involving the rs3512 site using Target scan human v6.2 (https://www.targetscan.org/vert_61/). Evaluation of the expression levels of miR-124-3p was based on small RNA Expression Atlas (https://microrna-atlas.omics.kitchen/).

miRNA mimic and inhibitor for miR-124-3p

miRNA mimic (MC10060) and inhibitor for miR-124-3p (MH10060) were purchased from Invitrogen. The miRNA negative control (i.e., scramble; #4464058) and negative control for inhibitor (#446476) were also obtained from Invitrogen.

Immunoblot analysis

Cells were washed with ice-cold PBS and incubated in protease inhibitor (#11697498001, Roche) containing cell lysis buffer (#9803, Cell signaling). Protein lysates were separated on 4-12% Bis-Tris gels (Invitrogen) and transferred to PVDF membranes. Membranes were incubated with primary antibodies for FAN1 (internally developed) and Actin (A4700, Sigma-Aldrich) overnight at 4°C. After washing, the blots were probed with anti-sheep and anti-mouse IgG-HRP secondary antibodies and visualized by ECL (PerkinElmer) as described previously (21). Densitometric quantification of bands were performed by Multi Gauge v2.3 software.

FAN1 3'-UTR reporter luciferase vectors

Total RNAs were isolated from a patient-derived LCL carrying 15AM2 and non-15AM2 *FAN1* using RNeasy mini kit (Qiagen). cDNA synthesis was performed by Superscript III first strand kit (Invitrogen), and then the full length 15AM2 and non-15AM2 3'-UTRs were amplified by Ex Taq DNA polymerase (Takara) using specific primers (5'-GGCTAGCAAGATTCCCTACA-3' and 5'-GCTCGAGTTAGTTTTGCAAATATTTA-3'). Subsequently, PCR products were digested by restriction enzymes, NheI and XhoI, and inserted into the pmirGLO dual luciferase miRNA target expression vector (Promega). To generate 3'-UTR luciferase reporter vectors with swapped alleles, site-directed mutagenesis was performed using Phusion high fidelity DNA polymerase (NEB) with original 15AM2 and non-15AM2 full length 3'-UTR reporter vectors as templates. To generate non-15AM2 carrying the alternative allele at rs3512 site (namely, non-15AM2-C), the following primer set were used: 5'-CAGTGTTACTCTTGGAACTGCCTTTACTTTTAACG-3' and 5'-CGTTAAAAGTAAAGGCAGTTCCAAGAGTAACACTG-3'. To generate 15AM2 carrying the reference allele at rs3512 (namely 15AM2-G), the following primers were used: 5'-CAGTGTTACTCTTGGAAGTGCCTTTACTTTTAACGC-3 and 5'-GCGTTAAAAGTAAAGGCACTTCCAAGAGTAACACTG-3'). The sequence of all constructs was confirmed by Sanger sequencing at the Massachusetts General Hospital Genomics core (Boston, MA).

mers were used: 5'-CAGTGTTACTCTTGGAAGTGCCTTTACTTTTAACGC-3 and 5'-GCGTTAAAAGTAAAGGCACTTCCAAGAGTAACACTG-3'). The sequence of all constructs was confirmed by Sanger sequencing at the Massachusetts General Hospital Genomics core (Boston, MA).

Dual luciferase assay to determine the 3'-UTR reporter gene activity

To measure the 3'-UTR luciferase activity, *FAN1* knockout HEK293T cells (21) were aliquoted in 96-well plates and co-transfected with 15AM2 or non-15AM2 full length 3'-UTR luciferase reporter vector and miR-124-3p (50 nM) with or without miR-124-3p inhibitor (100 nM) using Lipofectamine 2000 (24 hours). The procedure for 3'-UTR luciferase activity assay was the same

as the one used for luciferase promoter activity assay. To calculate normalized 3'-UTR reporter activity, the firefly luciferase signal (i.e., 3'-UTR activity) was divided by the corresponding Renilla luciferase signal (i.e., transfection control) and then, the luciferase signal of a test sample was normalized by negative control miRNA sample to generate relative levels.
(i.e., transfection control) and then, the luciferase signal of a test sample was normalized by negative control miRNA sample to generate relative levels.

Quantification of the CAG repeat instability using an expansion index method

The impacts of miR-124-3p on CAG repeat instability was determined *in vitro* due to 1) low sequence homology between mouse and human *FAN1* 3'UTR and 2) the lack of a mouse model carrying *Fan1* with human 3'-UTR sequence. To determine the impacts of miR-124-3p on the CAG repeat instability in cells, we analyzed patient-derived NPC with 130 CAGs because repeat expansion was generated in cells carrying a very long CAG repeat (20). Due to limited availability of patient-derived cells with long CAG repeat, we analyzed HD NPC cells from a patient-derived iPSC cells (ND50034; 130/22 CAGs;) carrying two copies of alternative alleles at rs3512. NPC cells were transfected weekly with either miR-124-3p or a scramble (100 nM) using human stem cell nucleofector kit 1 (VPH-5012, Lonza) for 5 weeks. Before each treatment, half of the cells were spared for repeat instability assay. The levels of CAG repeat expansion were determined as described previously (62). Briefly, genomic DNA was amplified by Qiagen Taq Core kit (#201225, Qiagen) with specific primers (5'-FAM- ATGAAGGCCTTCGAGTCCCTCAAGTCCTTC-3' and 5'-GGCGGCTGAGGAAGCTGAGGA-3'), and subjected to fragment analysis using ABI3730. The quantification of DNA fragments was done by Massachusetts General Hospital Genomics core (Boston, MA). Subsequently, we calculated the expansion index using a relative threshold of 5% as previously described (21).

Quantitative real time PCR (qRT-PCR)

The effects of actinomycyn D were determined by qRT-PCR. cDNA was synthesized from total RNA using Superscript III first strand kit, and subjected to qRT-PCR using the LightCycler 480 system (Roche) and LightCyler 480 SYBR Green I master (Roche, 04707516001). The following primer sets were used: *FAN1* (5'-TGGAAGTTAAAGGCCCCAAT-3', 5'-TGGCTCTTAGCTCCAACTGC-3'); *ACTIN* (5'-TCCCTGGAGAAGAGCTACGA-3', 5'-AGCACTGTGTTGGCGTACAG-3'). The Ct value of *FAN1* was divided by that of actin, further normalized by control sample (i.e., 2(-Delta Delta C(T)) (63).

Statistical analysis and software

All data are presented as mean ± standard errors of independent experiments. For group comparisons, Student t-tests were performed for group comparisons. Statistical analysis of RNAseq data was performed using generalized linear regression analysis. Multiple test correction was performed using false discovery rate using R 3.5.3. R 3.5.3 was also used to produce plots.

Acknowledgments

Supported by NIH grants NS091161, NS125742, NS049206, NS105709, and NS119471. This manuscript is dedicated to the late Dr. Lesley Jones, who had made invaluable contributions to this study and countless discoveries in Huntington's disease.

References

1. H. T. Orr, H. Y. Zoghbi, Trinucleotide repeat disorders. *Annual review of neuroscience* **30**, 575-621 (2007).
2. C. E. Pearson, K. Nichol Edamura, J. D. Cleary, Repeat instability: mechanisms of dynamic mutations. *Nature reviews. Genetics* **6**, 729-742 (2005).
3. HDCRG, A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell* **72**, 971-983 (1993).
4. G. P. Bates *et al.*, Huntington disease. *Nature reviews. Disease primers* **1**, 15005 (2015).
5. S. E. Andrew *et al.*, The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nature genetics* **4**, 398-403 (1993).
6. M. Duyao *et al.*, Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nature genetics* **4**, 387-392 (1993).
7. R. G. Snell *et al.*, Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. *Nature genetics* **4**, 393-397 (1993).
8. F. Persichetti *et al.*, Huntington's disease CAG trinucleotide repeats in pathologically confirmed post-mortem brains. *Neurobiology of disease* **1**, 159-166 (1994).
9. J. M. Lee *et al.*, CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion. *Neurology* **78**, 690-695 (2012).
10. GeM-HD Consortium, Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. *Cell* **162**, 516-526 (2015).
11. J. M. Lee *et al.*, A modifier of Huntington's disease onset at the MLH1 locus. *Human molecular genetics* **26**, 3859-3867 (2017).
12. GeM-HD Consortium, CAG Repeat Not Polyglutamine Length Determines Timing of Huntington's Disease Onset. *Cell* **178**, 887-900 e814 (2019).
13. M. Ciosi *et al.*, A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes. *EBioMedicine* **48**, 568-580 (2019).
14. M. Flower *et al.*, MSH3 modifies somatic instability and disease severity in Huntington's and myotonic dystrophy type 1. *Brain : a journal of neurology* **142**, 1876-1886 (2019).
15. D. J. H. Moss *et al.*, Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. *The Lancet. Neurology* **16**, 701-711 (2017).
16. J. M. Loupe *et al.*, Promotion of somatic CAG repeat expansion by Fan1 knock-out in Huntington's disease knock-in mice is blocked by Mlh1 knock-out. *Human molecular genetics* **29**, 3044-3053 (2020).
17. R. Wang *et al.*, DNA repair. Mechanism of DNA interstrand cross-link processing by repair nuclease FAN1. *Science* **346**, 1127-1130 (2014).
18. A. Smogorzewska *et al.*, A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. *Molecular cell* **39**, 36-47 (2010).
19. J. M. Lee *et al.*, Genetic modifiers of Huntington disease differentially influence motor and cognitive domains. *American journal of human genetics* **109**, 885-899 (2022).

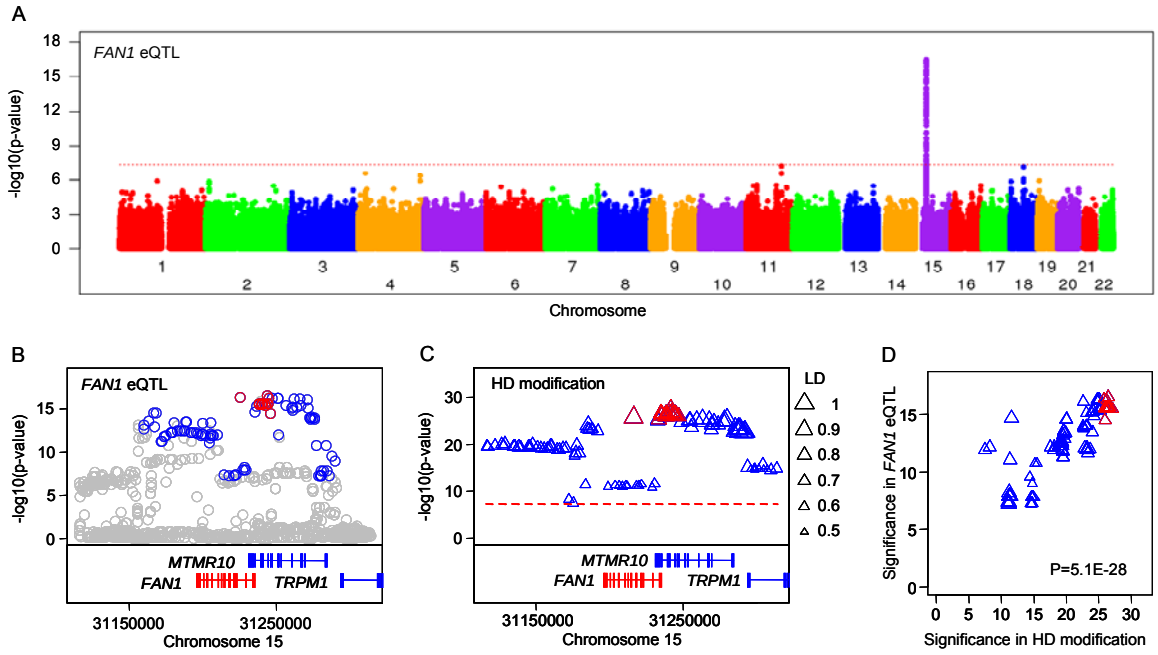
20. R. Goold *et al.*, FAN1 modifies Huntington's disease progression by stabilizing the expanded HTT CAG repeat. *Human molecular genetics* **28**, 650-661 (2019).
21. K. H. Kim *et al.*, Genetic and Functional Analyses Point to FAN1 as the Source of Multiple Huntington Disease Modifier Effects. *American journal of human genetics* **107**, 96-110 (2020).
22. B. McAllister *et al.*, Exome sequencing of individuals with Huntington's disease implicates FAN1 nuclease activity in slowing CAG expansion and disease onset. *Nature neuroscience* **25**, 446-457 (2022).
23. R. Goold *et al.*, FAN1 controls mismatch repair complex assembly via MLH1 retention to stabilize CAG repeat expansion in Huntington's disease. *Cell reports* **36**, 109649 (2021).
24. A. Porro *et al.*, FAN1-MLH1 interaction affects repair of DNA interstrand cross-links and slipped-CAG/CTG repeats. *Science advances* **7** (2021).
25. A. S. Phadte *et al.*, FAN1 removes triplet repeat extrusions via a PCNA- and RFC-dependent mechanism. *Proceedings of the National Academy of Sciences of the United States of America* **120**, e2302103120 (2023).
26. Y. Zou, P. Carbonetto, G. Wang, M. Stephens, Fine-mapping from summary data with the "Sum of Single Effects" model. *PLoS genetics* **18**, e1010299 (2022).
27. S. E. Castel, A. Levy-Moonshine, P. Mohammadi, E. Banks, T. Lappalainen, Tools and best practices for data processing in allelic expression analysis. *Genome biology* **16**, 195 (2015).
28. T. Qi *et al.*, Genetic control of RNA splicing and its distinct role in complex trait variation. *Nature genetics* **54**, 1355-1363 (2022).
29. P. Naeli, T. Winter, A. P. Hackett, L. Alboushi, S. M. Jafarnejad, The intricate balance between microRNA-induced mRNA decay and translational repression. *The FEBS journal* **290**, 2508-2524 (2023).
30. M. R. Fabian, N. Sonenberg, W. Filipowicz, Regulation of mRNA translation and stability by microRNAs. *Annual review of biochemistry* **79**, 351-379 (2010).
31. V. Agarwal, G. W. Bell, J. W. Nam, D. P. Bartel, Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4** (2015).
32. S. Grosswendt *et al.*, Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. *Molecular cell* **54**, 1042-1054 (2014).
33. N. S. Abell *et al.*, Multiple causal variants underlie genetic associations in humans. *Science* **375**, 1247-1254 (2022).
34. N. Segui *et al.*, Germline Mutations in FAN1 Cause Hereditary Colorectal Cancer by Impairing DNA Repair. *Gastroenterology* **149**, 563-566 (2015).
35. W. Zhou *et al.*, FAN1 mutations cause karyomegalic interstitial nephritis, linking chronic kidney failure to defective DNA damage repair. *Nature genetics* **44**, 910-915 (2012).
36. C. Bettencourt *et al.*, DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. *Annals of neurology* **79**, 983-990 (2016).
37. B. H. Laabs *et al.*, Identifying genetic modifiers of age-associated penetrance in X-linked dystonia-parkinsonism. *Nature communications* **12**, 3216 (2021).
38. X. N. Zhao, K. Usdin, FAN1 protects against repeat expansions in a Fragile X mouse model. *DNA repair* **69**, 1-5 (2018).
39. M. Lagos-Quintana *et al.*, Identification of tissue-specific microRNAs from mouse. *Current biology : CB* **12**, 735-739 (2002).

40. A. M. Krichevsky, K. S. King, C. P. Donahue, K. Khrapko, K. S. Kosik, A microRNA array reveals extensive regulation of microRNAs during brain development. *Rna* **9**, 1274-1281 (2003).
41. L. C. Cheng, E. Pastrana, M. Tavazoie, F. Doetsch, miR-124 regulates adult neurogenesis in the subventricular zone stem cell niche. *Nature neuroscience* **12**, 399-408 (2009).
42. F. B. Gao, Context-dependent functions of specific microRNAs in neuronal development. *Neural development* **5**, 25 (2010).
43. C. Saraiva *et al.*, MicroRNA-124-loaded nanoparticles increase survival and neuronal differentiation of neural stem cells in vitro but do not contribute to stroke outcome in vivo. *PloS one* **13**, e0193609 (2018).
44. S. Hamzei Taj *et al.*, Dynamic Modulation of Microglia/Macrophage Polarization by miR-124 after Focal Cerebral Ischemia. *Journal of neuroimmune pharmacology : the official journal of the Society on NeuroImmune Pharmacology* **11**, 733-748 (2016).
45. T. H. Rainer *et al.*, Plasma miR-124-3p and miR-16 concentrations as prognostic markers in acute stroke. *Clinical biochemistry* **49**, 663-668 (2016).
46. T. R. Doeppner *et al.*, MicroRNA-124 protects against focal cerebral ischemia via mechanisms involving Usp14-dependent REST degradation. *Acta neuropathologica* **126**, 251-265 (2013).
47. X. Ge *et al.*, Increased Microglial Exosomal miR-124-3p Alleviates Neurodegeneration and Improves Cognitive Outcome after rmTBI. *Molecular therapy : the journal of the American Society of Gene Therapy* **28**, 503-522 (2020).
48. S. T. Lee *et al.*, Altered microRNA regulation in Huntington's disease models. *Experimental neurology* **227**, 172-179 (2011).
49. E. Marti *et al.*, A myriad of miRNA variants in control and Huntington's disease brain regions detected by massively parallel sequencing. *Nucleic acids research* **38**, 7219-7235 (2010).
50. R. Johnson, N. J. Buckley, Gene dysregulation in Huntington's disease: REST, microRNAs and beyond. *Neuromolecular medicine* **11**, 183-199 (2009).
51. A. N. Packer, Y. Xing, S. Q. Harper, L. Jones, B. L. Davidson, The bifunctional microRNA miR-9/miR-9* regulates REST and CoREST and is downregulated in Huntington's disease. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **28**, 14341-14346 (2008).
52. R. Johnson *et al.*, A microRNA-based gene dysregulation pathway in Huntington's disease. *Neurobiology of disease* **29**, 438-445 (2008).
53. S. T. Lee *et al.*, Exosome-Based Delivery of miR-124 in a Huntington's Disease Model. *Journal of movement disorders* **10**, 45-52 (2017).
54. R. G. M. Relations (2021) Roche provides update on tominersen programme in manifest Huntington’s disease.
55. C. Sheridan, Questions swirl around failures of disease-modifying Huntington's drugs. *Nature biotechnology* **39**, 650-652 (2021).
56. R. Lewis (2021) Huntington’s Disease: 2 Steps Backward, 1 Step Forward.
57. M. R. Nelson *et al.*, The support of human genetic evidence for approved drug indications. *Nature genetics* **47**, 856-860 (2015).
58. Y. Zhang *et al.*, A genetic variant within MDM4 3'UTR miRNA binding site is associated with HPV16-positive tumors and survival of oropharyngeal cancer. *Molecular carcinogenesis* **58**, 2276-2285 (2019).

59. S. Bandiera *et al.*, Genetic variations creating microRNA target sites in the FXN 3'-UTR affect frataxin expression in Friedreich ataxia. *PloS one* **8**, e54791 (2013).
60. E. P. Hong *et al.*, Huntington's Disease Pathogenesis: Two Sequential Components. *Journal of Huntington's disease* **10**, 35-51 (2021).
61. B. van de Geijn, G. McVicker, Y. Gilad, J. K. Pritchard, WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nature methods* **12**, 1061-1063 (2015).
62. J. M. Lee *et al.*, A novel approach to investigate tissue-specific trinucleotide repeat instability. *BMC systems biology* **4**, 29 (2010).
63. K. J. Livak, T. D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408 (2001).

Figures

Figure 1. The most significant HD modifier *FAN1* haplotype represents *FAN1* eQTL.

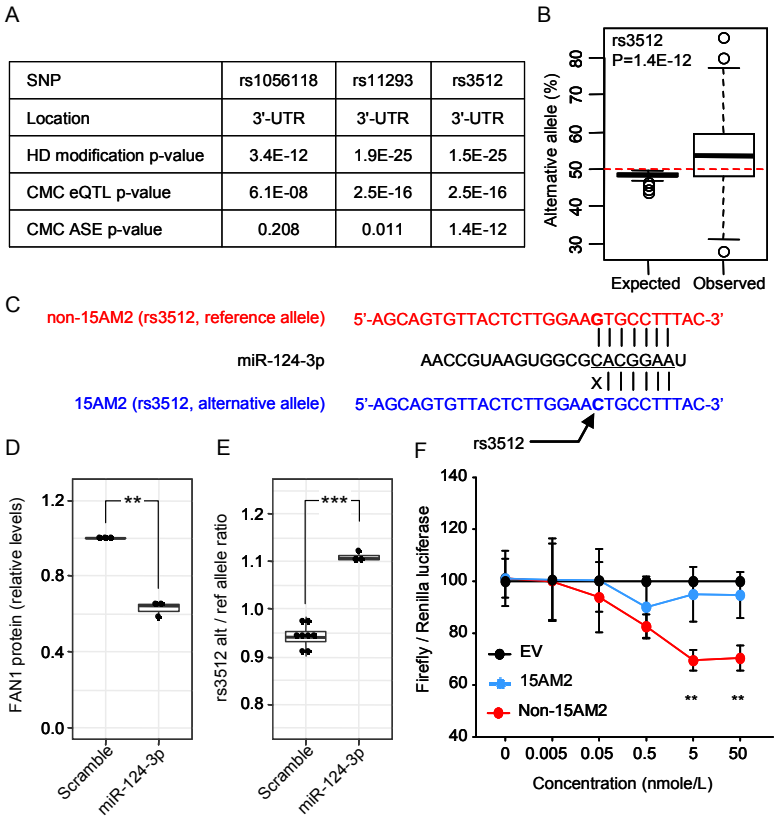


A) The results of a genome-wide eQTL analysis for *FAN1*, utilizing CMC data (prefrontal cortex, n=451), are plotted. SNPs with MAF > 1% are displayed. X-axis and Y-axis represents chromosomal location and -log10(P-value), respectively. B) *FAN1* eQTL signals at the chr15 region are summarized. Grey symbols represent non-tagging SNPs. C) HD modification signals marked by tagging (blue) and credible set SNPs (red) are plotted. The size triangle is proportional to LD. D) HD onset modification signals (X-axis; -log10(p-value)) of 15AM2-tagging SNPs were compared to the CMC *FAN1* eQTL signals (Y-axis; -log10(p-value)) for 15AM2-tagging SNPs. P-value was based on Pearson's correlation analysis. Blue and red symbols represent 15AM2-tagging SNPs and credible set SNPs, respectively.

Figure 2. rs3512 underlies HD modification, *FAN1* eQTL, and allelic imbalance.

A

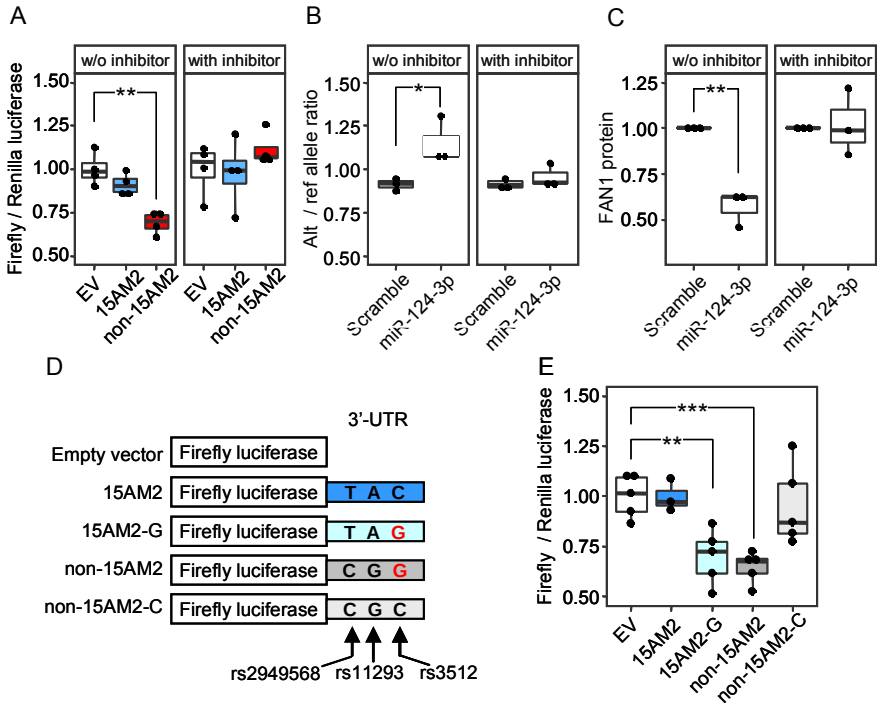
SNP	rs1056118	rs11293	rs3512
Location	3'-UTR	3'-UTR	3'-UTR
HD modification p-value	3.4E-12	1.9E-25	1.5E-25
CMC eQTL p-value	6.1E-08	2.5E-16	2.5E-16
CMC ASE p-value	0.208	0.011	1.4E-12



A) The significance levels of HD modification, *FAN1* eQTL, and allelic imbalance of three exonic 15AM2-tagging SNPs are summarized. B) The significance of allelic imbalance was determined by comparing observed ratio of alternative and reference alleles to null ratio using the CMC data (n=98). C) The TargetScan algorithm based on the reference sequence predicted potential interactions of miR-124-3p with *FAN1* 3'-UTR involving rs3512. Sequences in blue and red represent non-15AM2 and 15AM2 haplotypes carrying the reference and alternative alleles at rs3512, respectively. X represents a mismatch. D) HEK293T cells were treated with miR-124-3p or scrambled miRNA and analyzed by immunoblot assays to determine the impacts of the miRNA on endogenous FAN1 protein levels. E) MiSeq analysis of cDNA samples from an HD patient-derived LCL carrying heterozygous genotype at rs3512 was performed to determine the effects of miR-124-3p on allelic expression. Y-axis represents the ratio of the alternative allele to the reference allele at rs3512. F) *FAN1* knockout HEK293T cells were transfected with 3'-UTR reporter vectors and treated with various concentrations of miR-124-3p to determine the dose response relationship. The activity of 3'-UTR was normalized by the Renilla luciferase (Y-axis). Black, blue, and red circles represent empty vectors, 15AM2 3'-UTR, and non-15AM2 3'-UTR

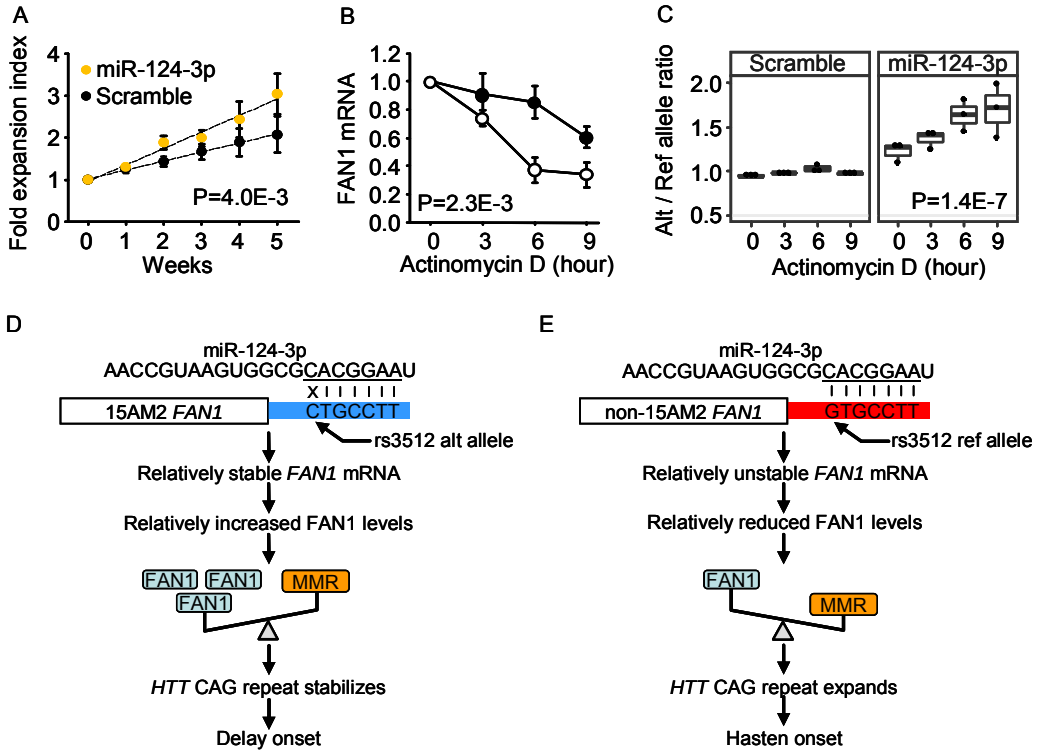
reporter vectors, respectively. Mean and standard errors are displayed (n = 3). **, p-value < 0.01; ***, p-value < 0.001 based on multiple test correction.

Figure 3. miR-124-3p preferentially targets the reference allele at rs3512.



A) *FAN1* knockout HEK293T cells were transfected with 3'-UTR reporter vectors and treated with miR-124-3p (50 nM) in the presence (right panel) or absence (left panel) of an antagomir for miR-124-3p (100 nM). 24hr post-treatment, firefly and Renilla luciferase activities were determined. EV, empty vector. B and C) An HD NPC line heterozygous at rs3512 was treated with miR-124-3p (100 nM) in the presence (left panel) or absence (right panel) of an antagomir for miR-124-3p (100 nM) for 48 hours to determine allelic expression by MiSeq analysis (C; n = 3) and FAN1 protein levels by immunoblot analysis (D; n = 3). D) To determine whether the effects of miR-124-3p were specific to the rs3512 site, we performed site-directed mutagenesis for 3'-UTR reporter vectors, generating two additional reporter vectors such as 15AM2 with reference allele (15AM2-G) and non-15AM2 with the alternative allele at rs3512 (non-15AM2-C). E) *FAN1* knockout HEK293T cells were transfected with reporter vectors and treated with miR-124-3p (50 nM). Following treatment (24 hours), firefly and Renilla luciferase activities were determined. At least three independent experiments were performed. **, $p<0.01$ and ***, $p<0.001$ by Student t-test after multiple test correction.

Figure 4. Post-transcriptional regulation of *FAN1* by miR-124-3p increases repeat instability.



A) A patient-derived NPC line carrying 130/22 CAG was treated with scramble (black) or miR-124-3p (gold; 100nM) for 5 weeks to determine the effects on *HTT* CAG repeat instability. The levels of repeat instability were quantified by the expansion index method, and the time-course changes in expansion index were calculated by dividing each data point by corresponding week 0 value. The significance of miR-124-3p was based on a linear regression analysis modeling changes in expansion index as a function of miRNA treatment (p-value, 0.004) corrected for time (p-value, 1.8E-9) and experimental batch (p-value, 0.0019). B and C) To determine whether miR-124-3p altered the stability of *FAN1* mRNA, an HD patient-derived LCL line (43/18 CAGs) heterozygous at rs3512 was treated with actinomycin D (10 mg/ml) in the absence (scramble) or presence of miR-124-3p. Subsequently, the *FAN1* mRNA levels (B) and allele count at rs3512 (C) were determined by qRT-PCR and MiSeq sequencing, respectively. To determine the significance of miR-124-3p, *FAN1* mRNA levels (B) or allelic ratio (C) was modeled as a function of actinomycin D treatment time and treatment (scramble vs. miR-124-3p) in a regression

analysis. Data represent mean ± standard error (n = 3). D and E) Models of rs3512-mediated modification of HD are illustrated. rs3512 changes the target binding site for miR-124-3p in the *FAN1* 3'-UTR, resulting in one mismatch for 15AM2 haplotype. Therefore, miR-124-3p targets the 15AM2 haplotype carrying the alternative allele at rs3512 less efficiently (D) compared to the non-15AM2 haplotype carrying the reference allele (E), resulting in the relative increase of *FAN1* levels in the HD subjects carrying 15AM2. An increase in *FAN1* levels shifts the balance between FAN1 and the mismatch repair pathway toward FAN1 pathway, resulting in stabilization of the repeat and delay HD manifestation.

S. Table 1. Fine-mapping analysis of 15AM2 *FAN1* tagging SNPs.

Fine-mapping analysis (using susieR program) was performed focusing on 112 SNPs that tagged the 15AM2 *FAN1* haplotype. This analysis yielded a credible set of 17 SNPs with coverage of 0.9514, meaning the proportion of credible set that contain a true causal SNP. VEP, variant effect predictor.

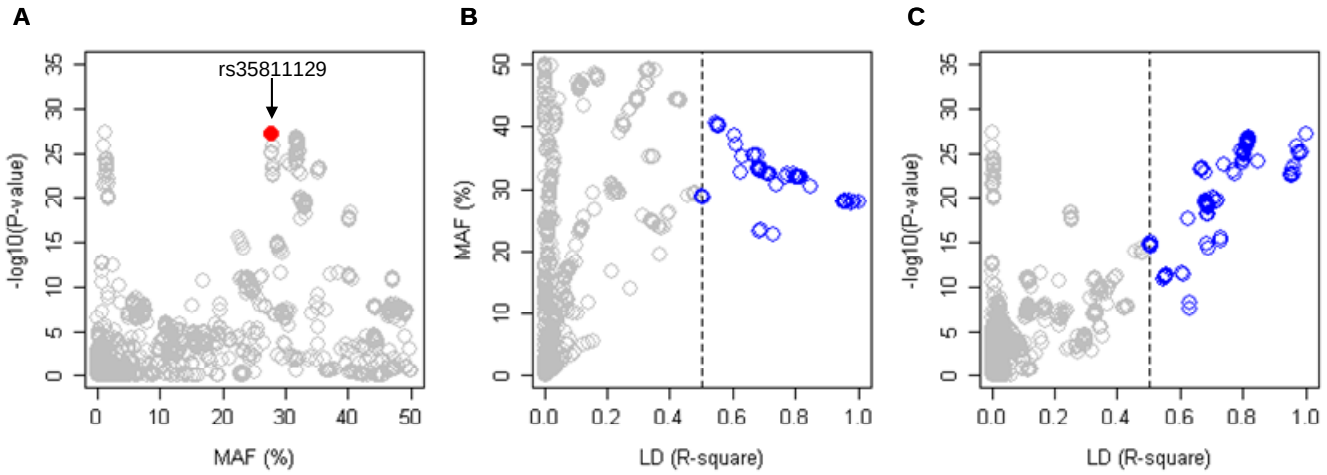
SNP	BP (hg19)	MAF (%)	VEP Annotation	Beta	P-value
rs61997076	31216225	28.1	intron_variant	0.18052	1.71E-26
rs3840035	31233352	32.3	3_prime_UTR_variant	0.171694	4.94E-26
rs11293	31234981	31.79	3_prime_UTR_variant	0.177055	2.75E-27
rs3512	31235005	31.8	3_prime_UTR_variant	0.177514	2.03E-27
rs7168641	31235452	31.88	intron_variant	0.175317	8.13E-27
rs35256769	31237776	31.92	intron_variant	0.174894	9.83E-27
rs61997079	31238271	31.86	intron_variant	0.176689	3.13E-27
rs11629793	31238423	31.93	3_prime_UTR_variant	0.174419	1.25E-26
rs11637469	31238783	31.88	3_prime_UTR_variant	0.176261	3.96E-27
rs8034856	31240957	31.88	intron_variant	0.177092	2.41E-27
rs35811129	31241346	27.81	intron_variant	0.186149	7.23E-28
rs762456260	31242461	31.84	intron_variant	0.178088	1.29E-27
rs11634761	31243173	31.87	intron_variant	0.176703	3.28E-27
rs2140734	31243792	31.87	intron_variant	0.176175	4.24E-27
rs7176569	31244495	31.87	intron_variant	0.176113	4.32E-27
rs8032942	31245175	31.85	intron_variant	0.175701	6.11E-27
rs34294266	31246840	31.96	intron_variant	0.174828	1.06E-26

S. Figure 1. Identification of the 15AM2-tagging SNPs based on the LD with the top SNP marking the frequent onset-delaying effect of *FAN1*.

SNPs tagging the effect of 15AM2 were identified based on the LD with the top SNP (rs35811129) that marked the common onset-delaying effect.

(A) The significance levels of SNPs in the chr15:31000000-31400000 region in our previous GWAS (Y-axis: -log10(p-value)) were compared to their minor allele frequencies (MAF) (X-axis), showing the 15AM2 top SNP rs35811129 that was analyzed as the lead SNP for LD calculation.

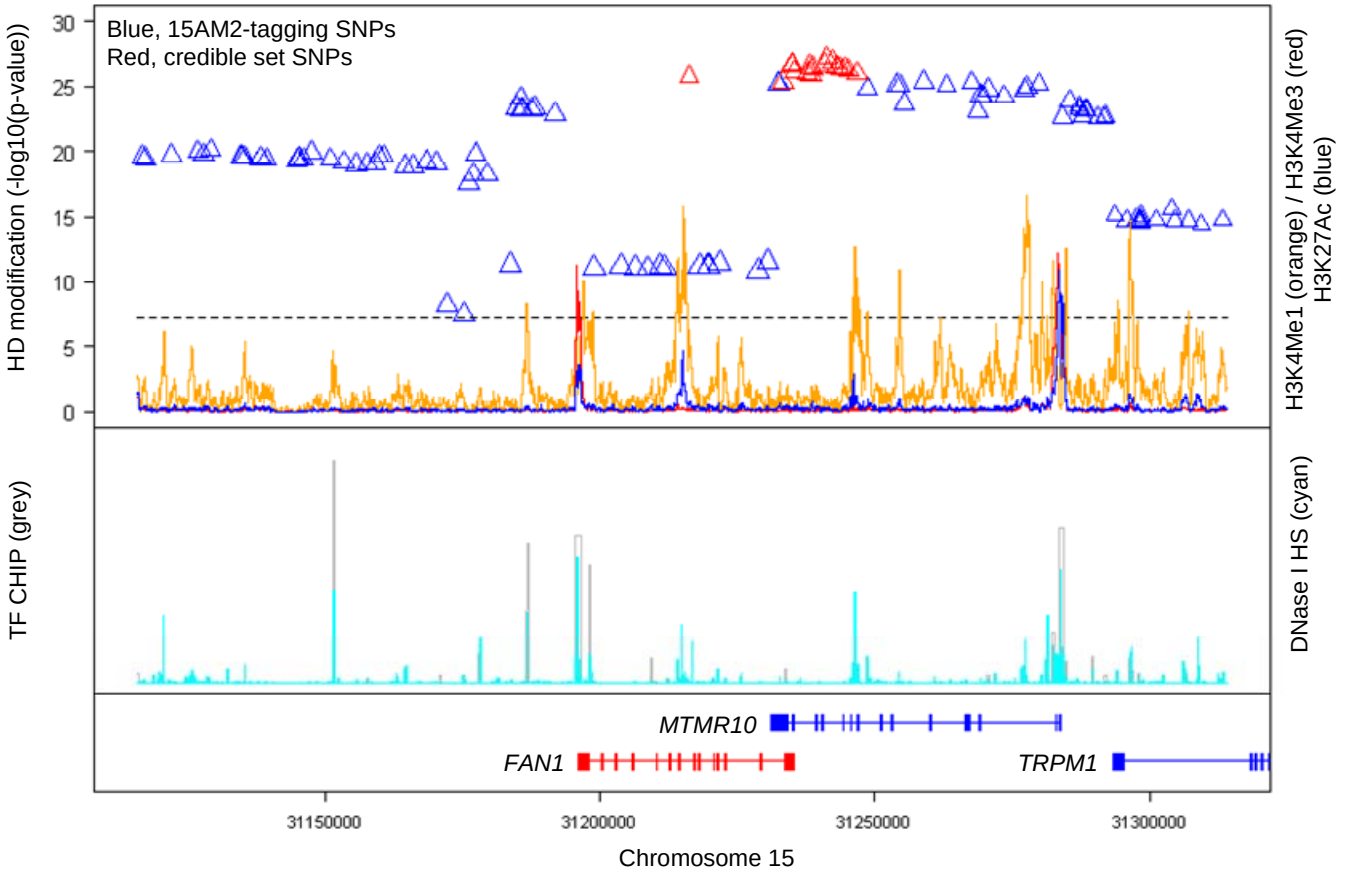
Subsequently, LD R-square values of SNPs with rs35811129 (X-axis) were compared to minor allele frequencies (MAF) (B; Y-axis) and association analysis p-values (C; Y-axis), revealing 112 15AM2-tagging SNPs (open blue circles) with LD R-square values > 0.5 (dashed black vertical lines in panels B and C).



S. Figure 2. The lack of correspondence between HD onset modification signals and ENCODE transcription regulation signals for 15AM2 modifier haplotype.

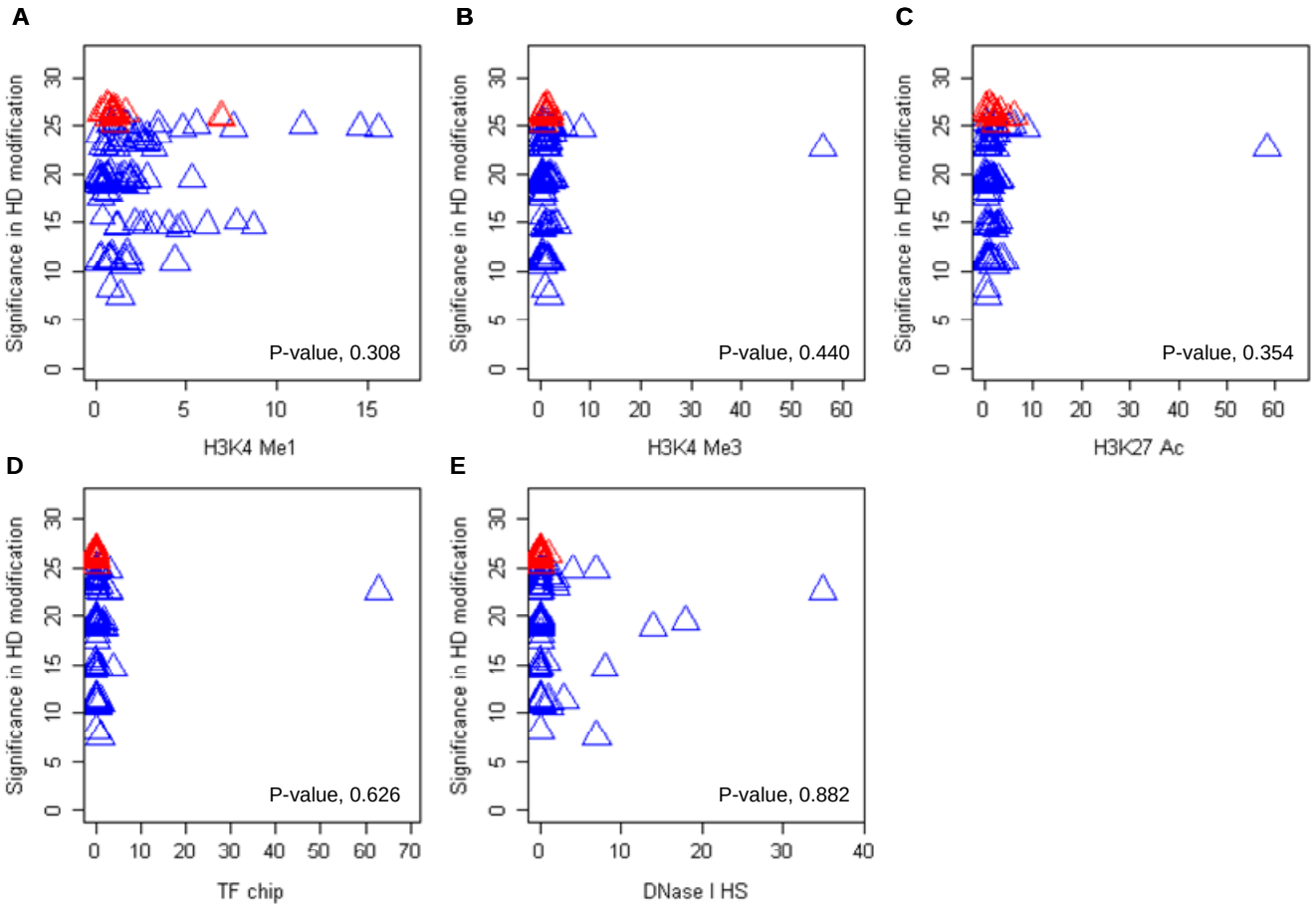
The 15AM2-tagging SNPs that captured the effect of frequent onset-delaying effects (primary Y-axis in the top panal) were compared to ENCODE transcription regulation signals (arbitrary unit). The size of the triangle is proportional to the MAF of the corresponding SNP. Traces in orange, red, blue, grey, and cyan represent H3K4Me1 (secondary Y-axis in the top panel, arbitrary unit), H3K4Me3 (secondary Y-axis in the top panel, arbitrary unit), H3K27Ac (secondary Y-axis in the top panel, arbitrary unit), TF CHIP (primary Y-axis in the middle panel, arbitrary unit), and DNase I hypersensitivity (secondary Y-axis in the middle panel, arbitrary unit), respectively. Blue and red triangles represent 15AM2-tagging SNPs and credible set SNPs.

IP (primary Y-axis in the middle panel, arbitrary unit), and DNase I hypersensitivity (secondary Y-axis in the middle panel, arbitrary unit), respectively. Blue and red triangles represent 15AM2-tagging SNPs and credible set SNPs.



S. Figure 3. The lack of correlation between onset modification signals and transcriptional regulation marks.

Focusing on 112 15AM2-tagging SNPs (blue) including 17 credible set SNPs (red), onset-modification signals (Y-axis; -log10(p-value)) were compared with H3K4Me1 (A), H3K4Me3 (B), H3K27Ac (C), TF CHIP (D), and DNase I hypersensitivity (E). P-values (bottom right corner) were based on Pearson's correlation analysis.

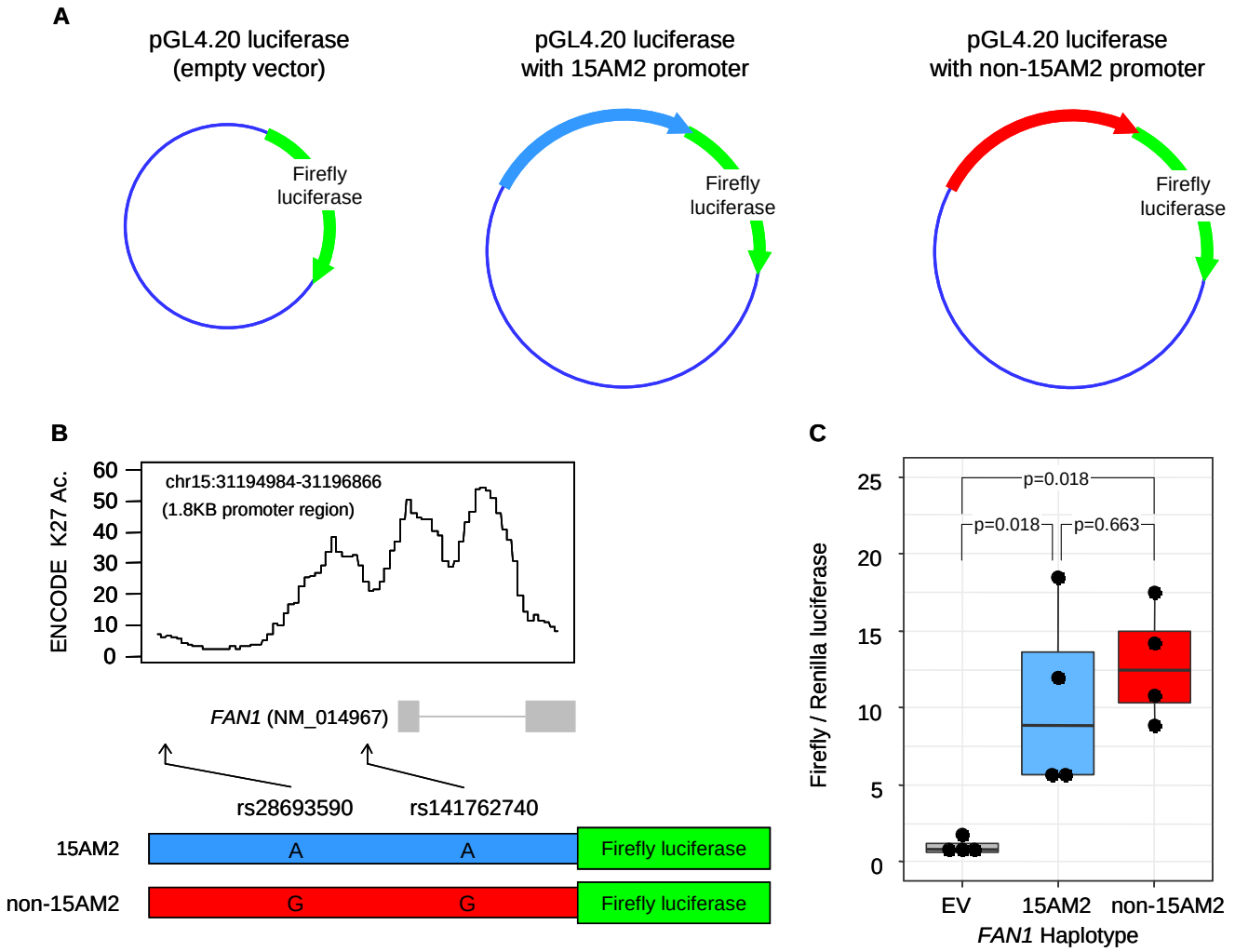


S. Figure 4. No difference in basal promoter activity between 15AM2 and non-15AM2.

(A) Firefly luciferase plasmids that were used to determine the promoter activity of 15AM2 and non-15AM2 are illustrated. 1.8 Kb promoter regions based on ENCODE K27 acetylation signals (chr15:31194984-31196866) were cloned into pGL4.20 firefly luciferase plasmid (empty vector; left) for 15AM2 (middle panel, blue) and non-15AM2 (right panel, red).

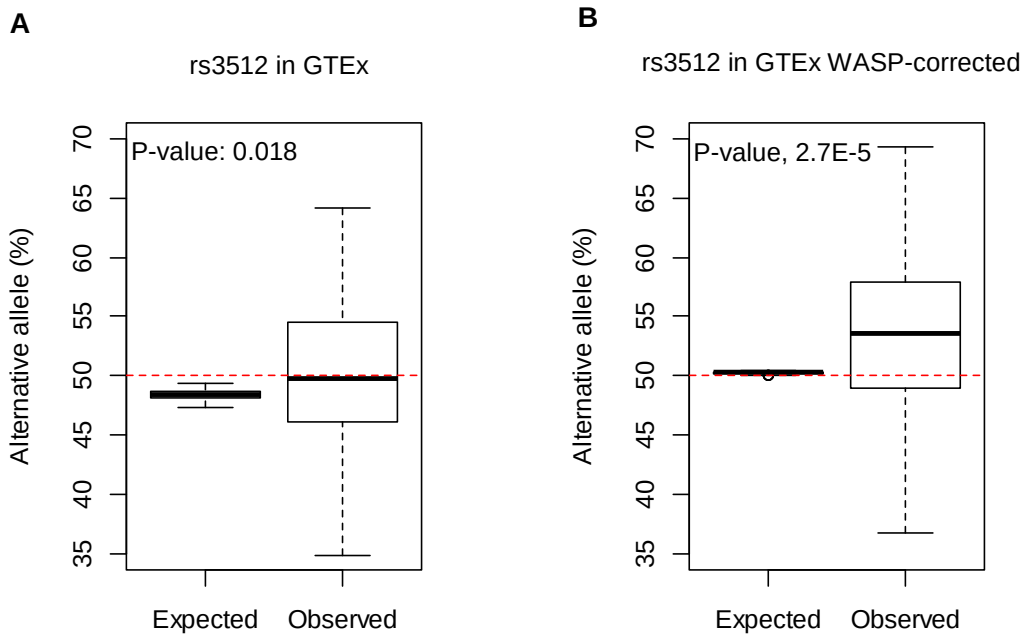
(B) Sanger sequencing analysis revealed that alleles at two SNP sites (rs28693590 and rs141762740) were different between 15AM2 (blue) and non-15AM2 (red) promoter regions (chr15:31194984-31196866). The ENCODE data are displayed at the top to show the locations of two SNPs relative to the three peaks of K27 acetylation signals.

(C) Firefly promoter reporter vectors were co-transfected with Renilla luciferase vector in HEK293 cells to determine the luciferase activities of empty vector (EV), 15AM2 (blue), and non-15AM2 (red) promoter region. Box plot shows the maximum, 75%, 50% (median), 25% quantile, and minimum. Adjusted p-values, based on the ANOVA analysis followed by TukeyHSD post-hoc test, are shown.



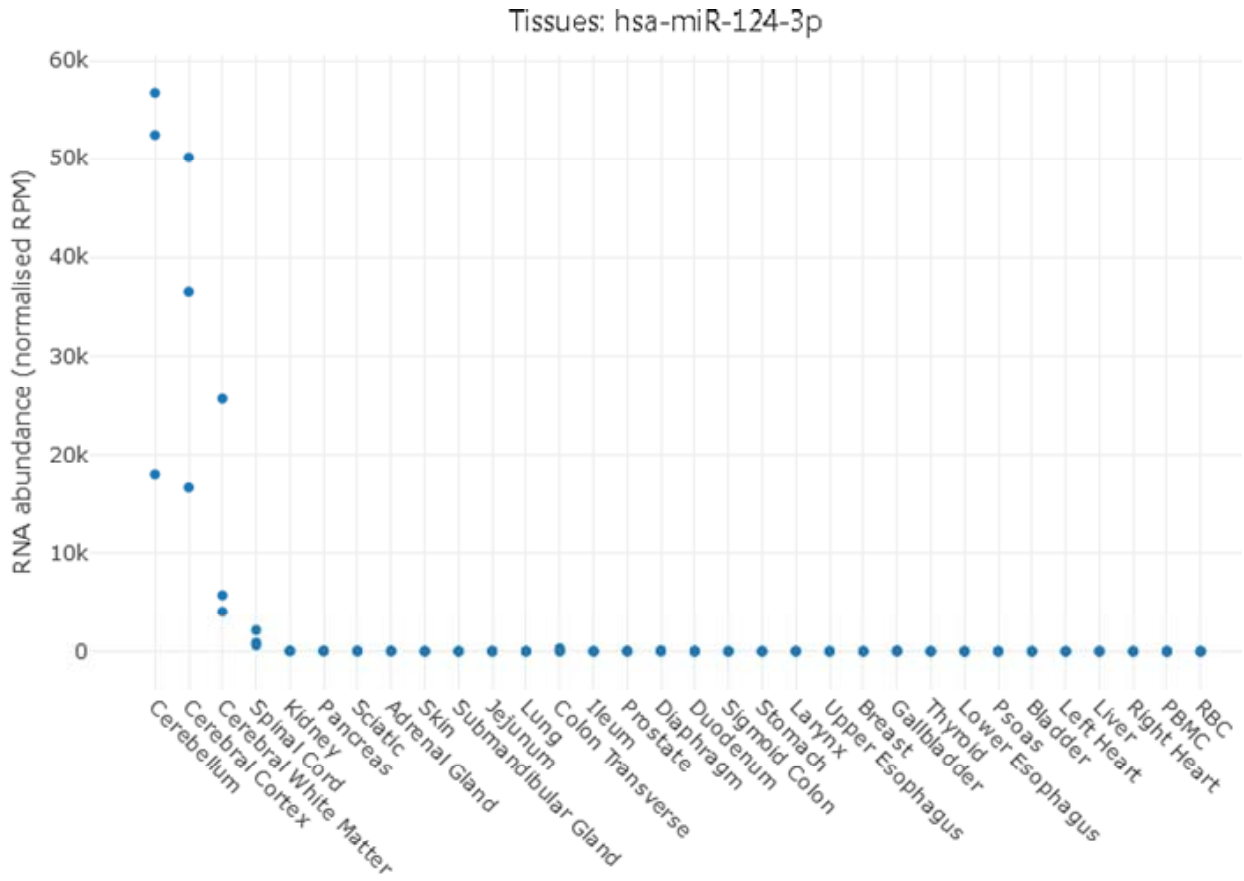
S. Figure 5. Significant over-representation of the alternative allele of rs3512 in bias-corrected ASE data.

(A) The observed proportions of observed alternative allele of rs3512 in heterozygous individuals in the GTEx (n=83) were compared to expected percentages of the alternative allele for C/G allele combination (p-value, 0.018). (B) We also compared the expected and observed percentages of the alternative allele at rs3512 in the 83 heterozygous individuals in the GTEx data using alignment bias-corrected ASE analysis, revealing a significant difference (p-value, 2.7E-5). Box plot shows the maximum, 75%, 50% (median), 25% quantile, and minimum.



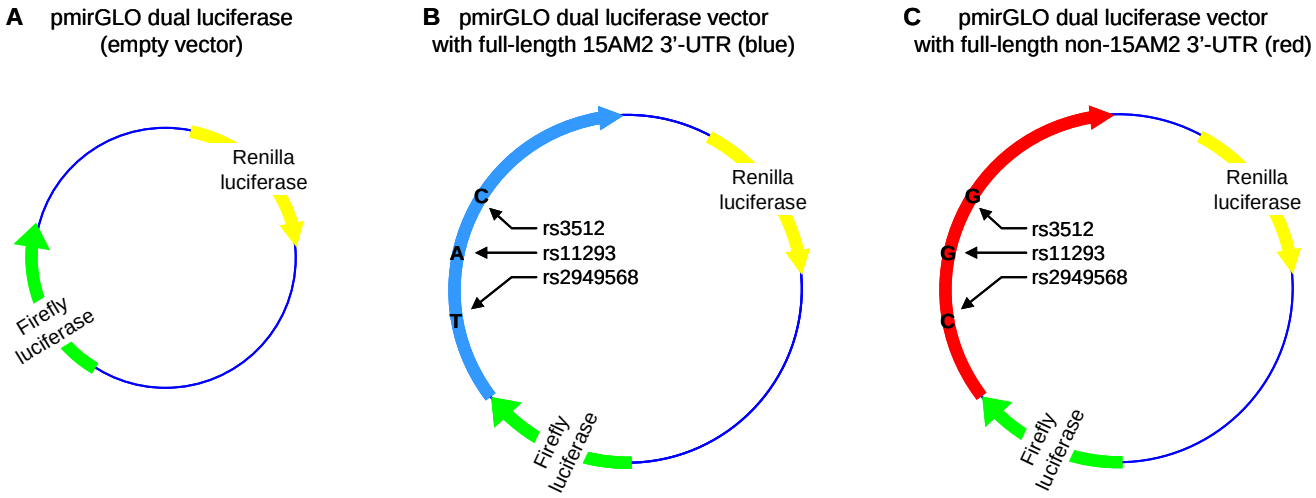
S. Figure 6. The expression levels of miR-124-3p in tissues.

The expression levels (RPM, reads per million) of miR-124-3p in human tissues were evaluated. The plot was captured from https://microrna-atlas.omics.kitchen/.



S. Figure 7. Full-length *FAN1* 3'-UTR reporter vectors to test the effects of candidate miRNAs.

Schematic diagrams of full-length *FAN1* 3'-UTR dual luciferase vectors for 15AM2 and non-15AM2 are illustrated (A, empty vector; B, 15AM2; C, non-15AM2). Alleles at three SNPs sites differ between 15AM2 and a representative non-15AM2 (rs2949568, rs11293, rs3512). The 15AM2 haplotype carries alternative alleles at all three SNP sites. In contrast, the non-15AM2 haplotype carries reference alleles.



S. Figure 8. Effects of 15AM3 on rs3512 in the association analysis.

(A) To determine whether the putative causal variant of 15AM3 (infrequent onset-hastening effect) contributed to this phenomenon, we constructed three SNP haplotypes comprising rs151322829 (putative causal variant of 15AM3), rs3512 (putative causal variant of 15AM2), and rs35811129 (top SNP tagging 15AM2). The yellow highlight represents minor alleles. Frequencies of haplotypes were based on our GWAS data.

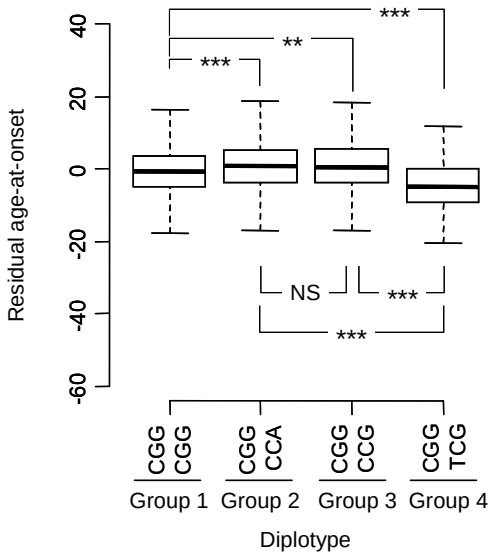
(B) Age-at-onset corrected for CAG repeat size (i.e., residual age-at-onset; Y-axis) of HD patients with different diplotypes are compared. All diplotypes carry one copy of the CGG haplotype as the baseline (the most frequent haplotype); Groups 1, 2, 3, and 4 carry CGG, CCA (15AM2), CCG (15AM2), and TCG (15AM3), respectively. **, multiple test-corrected p-value < 0.01; ***, multiple test-corrected p-value < 0.001.

(C) To confirm potential confounding by 15AM3 in the association analysis for rs3512, HD subjects without 15AM3 (analysis of groups 1 and 2 / analysis of groups 1, 2, and 3) and with 15AM3 (analysis of groups 1, 2, 3, and 4) were analyzed. Fixed effects model association analysis including a set of covariates that were used in our GWAS, was performed.

A

Haplotype	rs151322829 (Putative causal variant of 15AM3)	rs3512 (Putative causal variant of 15AM2)	rs35811129 (Top SNP tagging 15AM2)	Frequency (%)
CGG	C	G	G	68.70
CCA	C	C	A	27.15
CCG	C	C	G	3.24
TCG	T	C	G	0.85

B



C

Samples	P-value	
	rs3512	rs35811129
Groups 1 and 2 (analysis of full 15AM2 haplotype)	1.0E-15	1.0E-15
Groups 1, 2, and 3 (analysis w/o 15AM3)	7.5E-17	1.3E-13
Groups 1, 2, 3, and 4 (analysis with 15AM3)	9.5E-14	1.2E-15