

Supplemental Data

Five siRNAs Targeting Three SNPs May Provide Therapy for Three-Quarters of Huntington's Disease Patients

Edith L. Pfister, Lori Kennington, Juerg Straubhaar, Sujata Wagh, Wanzhou Liu, Marian DiFiglia, Bernhard Landwehrmeyer, Jean-Paul Vonsattel, Phillip D. Zamore, and Neil Aronin

Supplemental Experimental Procedures

Patient Samples, Sequencing, and Statistical Analysis

Patient brain samples were obtained from brain repositories in Charlestown, MA and New York, NY, and DNA from the DNA repository in Ulm, Germany. Genomic DNA was either extracted from brain tissue (USA patients) using a genomic DNA extraction kit (Lamda Biotech, St. Louis, MO) or obtained as purified (German patients). Candidate SNP sequences were amplified by PCR (Table S3) and sequenced (GENEWIZ, South Plainfield, NJ, USA, and Macrogen, Rockville, MD, USA). To identify new SNP sites, we selected six subjects for sequencing of all 67 *Huntingtin* exons. All electrophoretograms were manually inspected for the forward and reverse directions.

One hundred-nine case and 116 control genomes from German and US populations were typed at 24 SNP positions in the HD gene on human chromosome 4. Of these, 9 were discarded because they were rarely heterozygous. All assayed SNPs had a call rate greater than 95%. Four SNPs with a minimal allele frequency (MAF) of less than 0.01 were removed from the set. Deviations from Hardy-Weinberg equilibrium (HWE) were determined with Pearson goodness-of-fit and Fisher's exact tests. All markers resulted in HWE p -values of greater than 0.01. Single SNP associations were calculated for associations of markers with the HD phenotype. Test statistics of the

Pearson goodness-of-fit test was determined and significance evaluated against the chi-squared distribution and against an empirical distribution of the statistic after 1000 permutations. Association was also tested with the Fisher's exact test and the Cochran-Armitage test. A single marker, rs362307, was found to be associated with a significance of 0.0000523. This marker remained significant after Bonferroni multiple testing adjustment for 17 tests at the level of 0.000890. SNP rs362307 is located in a ~80 kb block of 10 markers whose average local linkage disequilibrium value is $D' = 0.995666$. The power of the study to detect association at $p < 0.01$ was $> 90\%$.

All statistical calculations were performed using the Haploview software, version 3.32 [S1] and R (R: Development core team (2004). R: A language and environment for statistical computing. Vienna, Austria. <http://www.r-project.org>). SNP data were imported into R and formatted for input into Haploview software.

Reporter Constructs and Assays

For rs363125, a 55-mer containing the SNP site (for 5'-cta gag GTT AAG AGA TGG GGA CAG TA[C/A] TTC AAC GCT AGA AGA ACA CAc tcg agc t-3', rev 5'-cta gag ctc gag TGT GTT CTT CTA GCG TTG AA[G/T] TAC TGT CCC CAT CTC TTA ACc t-3') was cloned into the pRL-TK vector (Promega Corporation, Madison, WI) using the XbaI site in the 3' UTR of the Renilla luciferase gene. Proper insertion was confirmed by PCR and sequencing. Luciferase assays were performed by co-transfection in 24 well plates of the siRNA with 0.025 ug/well of the SNP reporter (pRL3125) and 0.05 ug/well pGL3-control vector (Promega). For dose-response measurements, GFP siRNA (guide: 5'-GCA AGC UGA CCC UGA AGU UAA U-3'; passenger: 5'-GAA CUU CAG GGU CAG CUU GCC G-3') was added to each transfection mixture so that all transfections contained 20 nM total siRNA. Transfections were performed using Lipofectamine 2000 (Invitrogen Corporation, Carlsbad, CA), according to the manufacturer's protocol. Twenty-four hours after transfection the cells were lysed for 20 min in 1x passive lysis

buffer (Promega). Luciferase activity was read in 96-well plates with the Dual-luciferase assay kit (Promega) using the GloMax multi-detection system (Promega).

For rs362307 and rs362273, a 45-mer (rs362273 forward: 5'-tcg aAG CCA CGA G AA GCT GCT GCT [A/G]CA GAT CAA CCC CGA GCG GGA-3', reverse: 5'- ggc cTC CCG CTC GGG GTT GAT CTG [T/C]AG CAG CAG CTT CTC GTG GCT-3', rs362307 forward: 5'-tcg aCC GGA GCC TTT GGA AGT CTG [C/T]GC CCT TGT GCC CTG CCT CCA-3', reverse: 5'-ggc cTG GAG GCA GGG CAC AAG GGC [G/A]CA GAC TTC CAA AGG CTC CGG-3') containing the SNP site was cloned into pSiCHECK-2 (Promega) between the XhoI and NotI restriction sites in the 3' UTR of a codon-optimized form of the *Renilla reniformis* luciferase gene. We used 0.025 ug/well of the psiCHECK vector in our luciferase assays, which were performed as above. Data were graphed and analyzed using Igor Pro software (WaveMetrics, Portland, OR).

Western Blotting

Cells were grown and transfected in 6-well plates. The final concentration of total siRNA transfected in each well was 20 nM (GFP siRNA plus *Huntingtin* siRNA). An siRNA targeting a non-polymorphic site in the *Huntingtin* mRNA ("E1-4"; guide: 5'-UUC AUC AGC UUU UCC AGG GUC-3'; passenger: 5'-CCC UGG AAA AGC UGA UGA CGG-3') served as a positive control. Cells were lysed 48 h after transfection using Passive Lysis Buffer (Promega) supplemented with protease inhibitors (Roche Applied Science, Indianapolis, IN, USA). Samples were diluted in Laemmli Sample buffer (Bio-Rad Laboratories, Hercules, CA, USA) and resolved by electrophoresis through a 4–15% polyacrylamide denaturing Tris-HCl gel (Bio-Rad). After transfer to PVDF, blots were probed with anti-Huntingtin antibody (Ab1, 0.5 ug/ml) [S2] followed by an HRP-conjugated anti-rabbit secondary antibody (NA934V, GE Healthcare, Buckinghamshire, UK) diluted 1:10,000. Chemiluminescent detection was performed with SuperSignal West Dura Extended Duration Substrate (Thermo Scientific, Pierce, Rockford, IL, USA)

and images acquired with an LAS-3000 imaging system (Fujifilm, Tokyo, Japan). After probing with the anti-Huntingtin antibody, blots were stripped and re-probed with anti- α -Tubulin antibody (DM1A, Sigma Aldrich, St. Louis, MO, 1:1000) detected with anti-mouse secondary antibody (NA931V, GE Healthcare) diluted 1:10,000.

Quantitative PCR

Cells were grown and transfected in 6-well plates. The final concentration of total siRNA transfected in each well was 20 nM (GFP siRNA plus *Huntingtin* siRNA). RNA was extracted 24 h after transfection using TRI reagent solution (Ambion, Austin, TX), and then DNase treated with Turbo DNA-free DNase (Ambion). cDNA was synthesized using oligo(dT) primers, Superscript III reverse transcriptase (Invitrogen Corporation, Carlsbad, CA) and 0.5 μ g total RNA. Quantitative PCR reactions were performed with primers to amplify *Huntingtin* (forward, 5'-cgc aga gtc aga tgt cag ga-3'; reverse, 5'-ggg tct ctt gct tgt tcg ag-3') or β -actin mRNA (forward, 5'-gga ctt cga gca aga gat gg-3', reverse 5'-agc act gtg ttg gcg tac ag-3') using the Quantitect SYBR Green PCR kit (Qiagen, Valencia, CA). Data were analyzed using the $2^{-\Delta\Delta CT}$ method [S3] and β -actin mRNA for normalization.

Supplemental References

- S1. Barrett J.C., Fry B., Maller J., and Daly M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. *21*, 263-265.
- S2. DiFiglia M., Sapp E., Chase K., Schwarz C., Meloni A., Young C., Martin E., Vonsattel J.P., Carraway R., Reeves S.A. et al. (1995). Huntingtin is a cytoplasmic protein associated with vesicles in human and rat brain neurons. *Neuron*. *14*, 1075-1081.
- S3. Livak K.J., and Schmittgen T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-(\Delta\Delta C(T))}$ Method. *Methods*. *25*, 402-408.
- S4. DiFiglia M., Sena-Esteves M., Chase K., Sapp E., Pfister E., Sass M., Yoder J., Reeves P., Pandey R.K., Rajeev K.G. et al. (2007). Therapeutic silencing of

mutant huntingtin with siRNA attenuates striatal and cortical neuropathology and behavioral deficits. *Proc Natl Acad Sci U S A.* 104, 17204-17209.

- S5. Liu W., Kennington L.A., Rosas H.D., Hersch S., Cha J.H., Zamore P.D., and Aronin N. (2008). Linking SNPs to CAG repeat length in Huntington's disease patients. *Nat Methods.* 5, 951-953.

Supplemental Figure Legends

Figure S1. Repression of Luciferase Expression in Reporter Assays Corresponds to Depletion of Endogenous *Huntingtin* mRNA

(A) Matched and mismatched siRNAs targeting the SNP rs363125 at nt 5,304 of *Huntingtin* mRNA. siRNAs are shown in capital letters with the passenger strand at top and the guide stranded paired to the mRNA, in lower case letters. The mismatch is at siRNA position 10. Dose-response analysis for these siRNAs using (B) transfected plasmids expressing luciferase reporters and (C) quantitative RT-PCR assays measuring endogenous *Huntingtin* mRNA. HeLa cells are homozygous for the C isoform of this SNP (data not shown).

Figure S2. A Fully Matched siRNA That Reduces Expression of Both a Luciferase Reporter and Endogenous *Huntingtin* mRNA Causes a Corresponding Depletion of Endogenous Huntingtin Protein

An identical siRNA, but for a position 10 (P10) mismatch to the luciferase reporter and to the endogenous *Huntingtin* mRNA, was far less effective at suppressing Huntingtin protein production. HeLa cells were transfected with either the P10 match or P10 mismatch siRNA targeting SNP rs363125 at nt 5,304 of the *Huntingtin* mRNA, GFP siRNA alone, or a positive control siRNA targeting a non-polymorphic site in Exon 1 (E1-4) of the *Huntingtin* mRNA [S4]. Cells were lysed 48 h after transfection and analyzed by Western blotting using antibodies to *Huntingtin* and α -Tubulin, which served as a loading control. (A) and (B) show independent replicates of the experiment.

Figure S3. Representative Data for the Development of an Allele-Specific siRNA-Targeting SNP rs362307, Which Is Associated with HD

(A) siRNAs targeting the U isoform of rs362307 and mismatched to the C isoform at either position 10 or position 16 did not discriminate between matched and mismatched luciferase reporter mRNAs. (B) Placing an additional mismatch in the seed sequence of the siRNA bearing a position 10 mismatch to the C isoform improved its selective targeting of the U isoform. (C) A doubly mismatched siRNA targeting the C isoform also distinguished between reporter mRNAs corresponding to the position 10 matched, C isoform and the position 10 mismatched, U isoform.

Figure S4. Representative Data for the dsRNAs Targeting the rs363125 SNP Site

siRNAs bearing a mismatch at position 10 to the C isoform (A) or the A isoform (B) of rs363125 did not discriminate well between matched and mismatched targets.

Figure S5. Adding a Position 5 Mismatch to a Position 10 Mismatch Increased the Ability of an siRNA to Discriminate between the Two Isoforms of the rs362273 SNP

We evaluated the efficacy and selectivity of siRNAs combining mismatches at positions 2, 3, 4, 5, 6, or 7 with a mismatch at position 10. The position 10 + position 5 siRNA was best able to distinguish between the matched and mismatched reporters.

Representative data are shown.

Supplemental Tables

Table S1. The U Isoform of SNP rs362307 at *Huntingtin* mRNA Nucleotide 9,633 Is Associated with the Expanded CAG Disease Allele

Linkage of CAG repeat length and SNP isoform identity was determined using SLiC [S5] for SNP rs362307 (*Huntingtin* nucleotide 9,633) for 16 patient blood samples. Of the 16 samples, eight were heterozygous for the rs362307 SNP, and the U isoform was linked to the expanded CAG repeat for seven of the eight.

Patient number	Linkage		
	<u>Nucleotide</u>	<u>Mutant allele</u>	<u>Normal allele</u>
4	C/U	U	C
5	C/U	U	C
7	C/U	U	C
8	C/U	U	C
9	C/U	C	U
11	C/U	U	C
14	C/U	U	C
15	C/U	U	C

Table S2. Validation of siRNAs Designed to Discriminate between Isoforms of the rs362037 SNP
The IC50 is reported as > 20 nM for siRNAs that failed to achieve half maximal inhibition at the highest concentration tested.

siRNA guide strand	SNP position	primary mismatch	secondary mismatch position	secondary mismatch	IC50 (nM)		Discrimination ratio
					Match	Mismatch	
5'-uacagacuuccaaaggcuccg-3'	2	A:C	none	none	0.30 ± 0.07	0.23 ± 0.04	0.77
5'-ucaagacuuccaaaggcucc-3'	3	A:C	none	none	0.52 ± 0.10	0.94 ± 0.62	1.8
5'-ugcagacuuccaaaggcuc-3'	4	A:C	none	none	0.88 ± 0.32	6.0 ± 3.7	6.8
5'-uggcagacuuccaaaggcu-3'	5	A:C	none	none	0.66 ± 0.32	1.8 ± 0.33	2.7
5'-ugggacagacuuccaaaggc-3'	6	A:C	none	none	0.93 ± 0.29	2.6 ± 0.89	2.8
5'-uagggacagacuuccaaagg-3'	7	A:C	none	none	0.45 ± 0.09	0.88 ± 0.49	1.9
5'-uaagggacagacuuccaaag-3'	8	A:C	none	none	0.36 ± 0.11	0.53 ± 0.12	1.5
5'-ucaagggacagacuuccaaa-3'	9	A:C	none	none	1.07 ± 0.06	0.93 ± 0.27	0.87
5'-uggcacaagggcacagacuuc-3'	13	A:C	none	none	0.25 ± 0.10	0.42 ± 0.11	1.7
5'-guagggcacacaagggcacagac-3'	16	A:C	2	U:G	3.5 ± 2.9	>20	> 5.7
5'-gccgggcacacaagggcacagac-3'	16	A:C	3	C:U	>20	>20	~1
5'-gcauggcacacaagggcacagac-3'	16	A:C	4	U:C	>20	>20	~1
5'-gcagugcacacaagggcacagac-3'	16	A:C	5	U:C	18 ± 8	>20	> 1.1
5'-gcaggucacacaagggcacagac-3'	16	A:C	6	U:C	7.8 ± 5.5	>20	> 2.6
5'-gcaggguacacaagggcacagac-3'	16	A:C	7	U:G	0.74 ± 0.08	9.4 ± 3.9	12.7
5'-gcagggcacacaagggcacagac-3'	16	A:C	15	A:G	>20	>20	~1
5'-gcagggcacacaagggcacagac-3'	16	A:C	15	U:G	>20	>20	~1
5'-gcagggcacacaagggcauagac-3'	16	A:C	17	U:G	1.0 ± 0.12	5.1 ± 0.67	5.1
5'-gcagggcacacaagggcaagac-3'	16	A:C	17	A:G	> 20	> 20	~1
5'-cagggcacacaagggcauagacu-3'	15	A:C	16	U:G	0.38 ± 0.02	0.64 ± 0.06	1.7

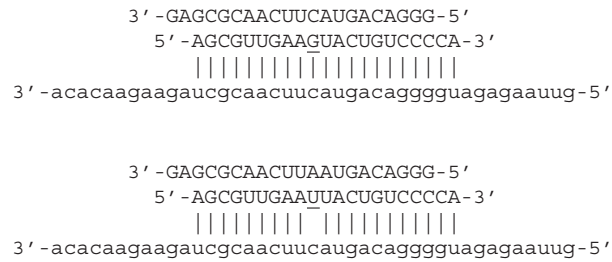
Table S3. Primers Used for SNP Analysis and Resequencing

Exon	SNP i.d.	Forward primer	Reverse primer
2		5'-TGGAGTGGGTAATTCAACACA-3'	5'-GCCAGAAATATGGGAAAAGG-3'
3		5'-AGAATTCCATGCAGGACACC-3'	5'-GCAGACATCCTCAGGGACTC-3'
4		5'-TGATGGGATGTGTCTTCCAT-3'	5'-GGTCAGGAGTTCGAGACCAG-3'
5		5'-ATGCAACCCTCTTGGTGACT-3'	5'-CGACAAAAACCAACATCCAG-3'
6		5'-TCAGCTGAGTTTTCCCATC-3'	5'-GAAGCACTCCCACAGGACTC-3'
7		5'-CTGCTCTTGAGTGCCCAA-3'	5'-CCACTCATATGCCTCCACCT-3'
8		5'-CTCTGGAAAGGACCTTGCTG-3'	5'-ATTCACATGCAGGGCCTAGA-3'
9		5'-TTGGTGGAAGTGATAGGAAA-3'	5'-GTTTTGGCAAGGAAGATGGA-3'
10		5'-TGCGATGTTAAGTGTTCCTG-3'	5'-CCTGGTTATCAGATTCCAGCA-3'
11		5'-GCATTTACTTAATTTGAAGTCCTTAT-3'	5'-CGAATATGCCCCATTTAAGC-3'
12		5'-CGTTATTTGCAAGCCTGTG-3'	5'-CTCCCAAAGTGCTGGGATTA-3'
13, 14		5'-GTTGGAGGGCTTGTCTCTTG-3'	5'-CAGGGATGGGAAAGCAATAA-3'
15, 16		5'-ACCTGGCTTAAGTGCTGCTC-3'	5'-CTCGGCTAGTGAAAACCAAA-3'
17, 18		5'-GTTCCATGGCTGAGCAATTT-3'	5'-GCTGAGAGATGGATACATGGTG-3'
19		5'-CTTGCCTTGGACCTTGTGTT-3'	5'-TGCATCAAGTGATCCCAGAA-3'
20, 21	rs363075	5'-CAAGCTGGCGTAAGTGTTT-3'	5'-TCCCTCTCTTCCATTCTCG-3'
22		5'-AAGTGGTGTCCGCTGGTAAC-3'	5'-GCCTAAAGAAAGGCATCAGG-3'
23		5'-CGTTTCACTTAAAAGTTGAGACTGC-3'	5'-TTTCTTAGCAAGCCTCATGGA-3'
24		5'-CTTTGTGGTGTGGGTGTG-3'	5'-TCCCACAGCTCCTGTCACTA-3'
25	rs35892913, rs1605746, rs17781557	5'-TGTGACATGCCTTCCTCTTG-3'	5'-AAAGGGACAAGCCATCACTG-3'
26		5'-CAGTTCCCAAGCAATTTGT-3'	5'-CCATCCACATGGTCACATTT-3'

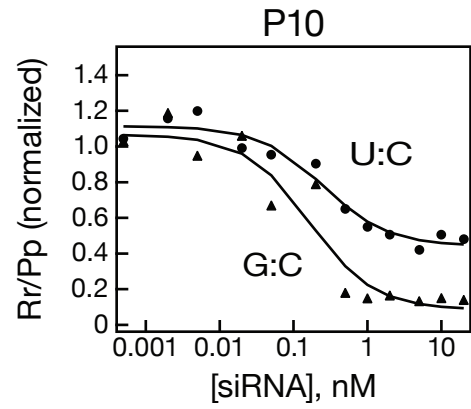
27		5'-TCAGGGTCCAAGAACAAAATG-3'	5'-GCTTCAGACCAAAGGTGGT-3'
28		5'-TTTCCAGTAATCTCTTTAAACTTGG-3'	5'-TAAAAGATGCAGAGGCCCAT-3'
29	rs4690074	5'-GGCCAGTAACCGTGTGTTCT-3'	5'-TCATGGCTAAGGCAGAGTCA-3'
30		5'-GGATTTCGTACAATAACGGGTCA-3'	5'-GGAGCTTCTGGTGTCTCTG-3'
31		5'-TTCACGCTGTGAGTCTTTGC-3'	5'-CTCTTTCGTGCTTCCACCA-3'
32		5'-TGCTTCCCTTTTATTCCATT-3'	5'-CCTGGAAAGTCTCAGCTCCA-3'
33		5'-TGCTTGAAGCTTTTAGTTGAAGG-3'	5'-ATGAGGGAAACATGCAGACC-3'
34		5'-TGTGAAATTTTATTTTCTTCCTG-3'	5'-TTCCATTTAAAGAAAACAGCAAAA-3'
35		5'-TGATGTGTGCTTGCTGTCAA-3'	5'-ACACACATGCAGAGCCTGAG-3'
36		5'-GATGTTGAGAGCAGTTTTCCAA-3'	5'-GCCCAAACCTGGTTCAAAGT-3'
37		5'-CGTCTCTTGGCAGCAGACTT-3'	5'-TATGCAACAACAAGCCAAGG-3'
38		5'-GGGTACAGGAAGCTGTCGTT-3'	5'-GCCCTACCCAACTGACTGA-3'
39	rs363125	5'-GCAATTGGGGGAAATTTAATC-3'	5'-CATCACGTGACTTCCCAAAA-3'
40		5'-TGTATACTTGGCGTAAGTGCTTT-3'	5'-ACTGGGCAAGGCAGAGTTT-3'
41		5'-GGACCGAGATGAAAGCAAAG-3'	5'-GCCAAAGCTCAGGTTACTGG-3'
42		5'-CTCACTGCCATCCAGAAACA-3'	5'-TTTAGTTTCGATGGAGCTTGG-3'
43		5'-GGCATTAAACCTGGTCTCTTCTT-3'	5'-TTTAAGGCAGGGAAAACACTGC-3'
44		5'-ATTGCCAGTTGCAGTTTTCC-3'	5'-AAAAGCCAGCCACCTGTTTT-3'
45		5'-TGAAGTGTACACATCAGTTCATCC-3'	5'-TAAACCCACCTATAAGGCACATC-3'
46		5'-TGTATTTTCTTTAAGAAGCCACT-3'	5'-ACAGGTGACAGAGGCACTCA-3'
47		5'-AGCTCCAGGGATGTGAAGTC-3'	5'-CAGACTGGAGTCCCCAACAT-3'
48	rs362336	5'-TGTTTGTTAACCTTTAATGCTCTGA-3'	5'-TATACTGGCCCTGGAATGCT-3'
49		5'-GCTTGACTGCCTTTCGAAGT-3'	5'-TGGAAAAGTGACTGGACTGG-3'
50	rs362331	5'-GGGCATTCTGTGACTCGGTA -3'	5'-GATAGGAACCCACCGTTCAT-3'
51		5'-GGCTAGTCTGTCTATCCCTTTCA-3'	5'-TCCAGGAGTCCCACTCACC-3'

52		5'-CAGCTGGTTGTAGGTCATGC-3'	5'-GGTCTTCTGCAAGGAACGAG-3'
53		5'-GCTTCCTGCTTCCTCACAGT-3'	5'-TTGCCAACACTGCAAATGT-3'
54		5'-ACAGGCTTGAGAAGGGTTGA-3'	5'-AGACCTCAGCAGGCTTTGTC-3'
55		5'-GAGGTGGTTGTGGGTGTCTT-3'	5'-CACCTTTGGGTCTGCATCTC-3'
56		5'-CACGGACAGGTGCTCACTTA-3'	5'-GGTGAGCATGCCAGTCTTCT-3'
57	rs362273	5'-AGTGACAAATCCCCAAGACC-3'	5'-GAGCTTTTCTCCTGGGTGTG-3'
58, 59		5'-TAGACGGTAGGCATGTGCTG-3'	5'-GTGTGGCCTGTGTGTGTGTT-3'
60		5'-GGATTCTAACAGCGGATTC-3'	5'-GTTCCGGGTCAACTCTTGAA-3'
61	rs362272	5'-CGGCCTGCTGTGTAGTCTCT-3'	5'-TCTTGCGTCTCACTGACCTC-3'
62, 63		5'-ACATGCTGTGAAGCCCTCTC-3'	5'-GTCGAGGTCCCTTGAGTGAG-3'
64		5'-CCCCTGTGTACAAAGCACTG-3'	5'-GCTGTGGTGGGGAATCACT-3'
65	rs3025806	5'-ATTTACATCGGCATTTTCC-3'	5'-AACTCCACCTCCAGGCTTTC-3'
66		5'-GAAAGCCTGGAGGTGGAGTT-3'	5'-ACATGAGCCTCGGTGTTGAC-3'
67 (primer set 1)	rs362308, rs362307	5'-GCTCTGCTCGCTCTCCAG-3'	5'-GCAGAGACACGCACGTTG-3'
67 (primer set 2)	rs362306, rs362268, rs362305, rs362304, rs362303	5'-TGACCAGGTCCTTCTCCTG-3'	5'-GGCCTTGCGATTCACATACT-3'
67 (primer set 3)	rs1557210, rs362302, rs3025805, rs362267	5'-ATGGATGCATGCCCTAAGAG-3'	5'-TCTAGGGCTGAGGAAGCAGA-3'

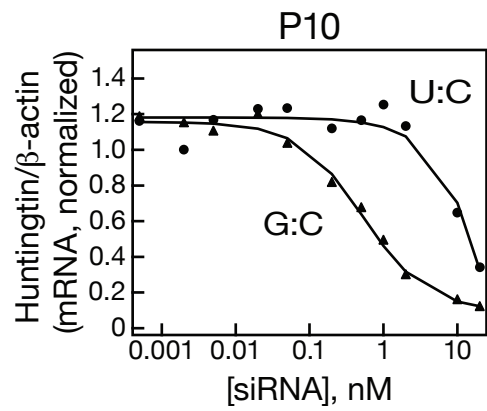
A

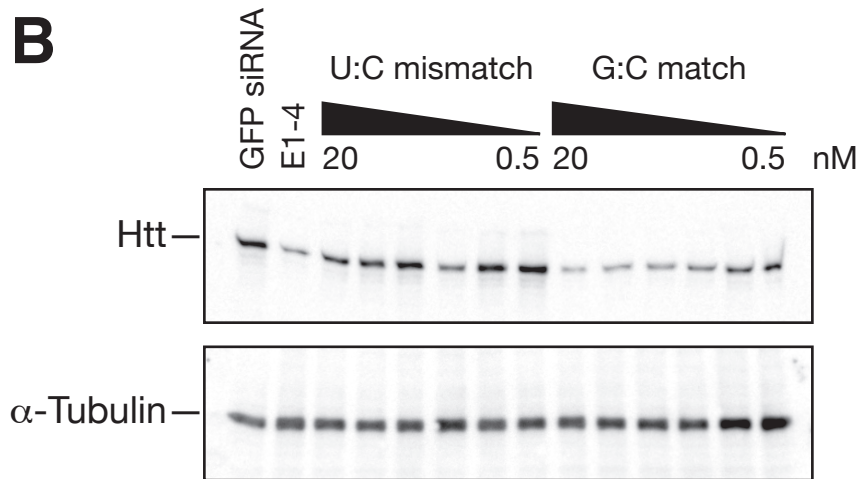
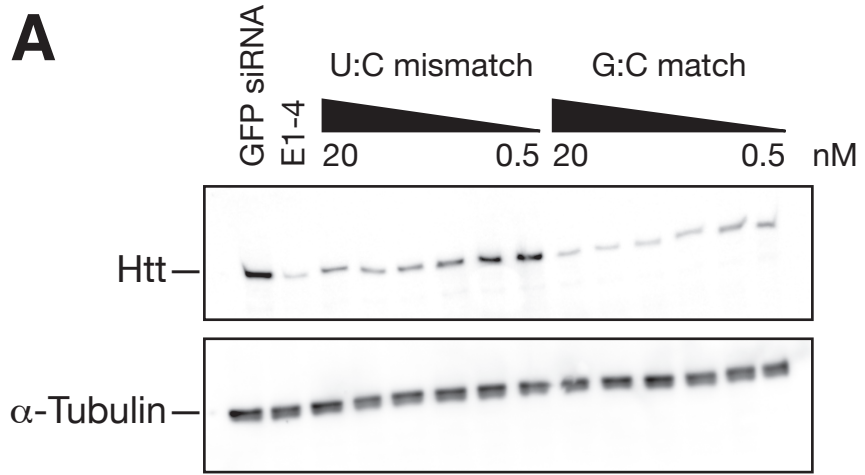


B

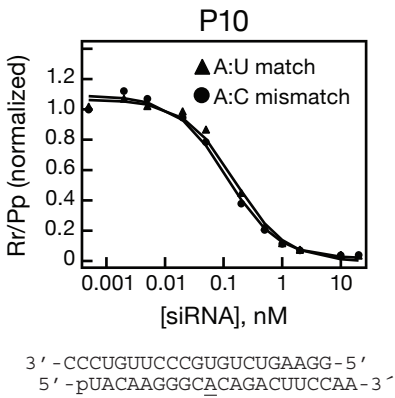


C

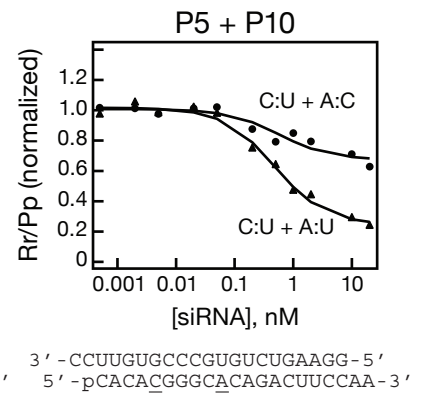
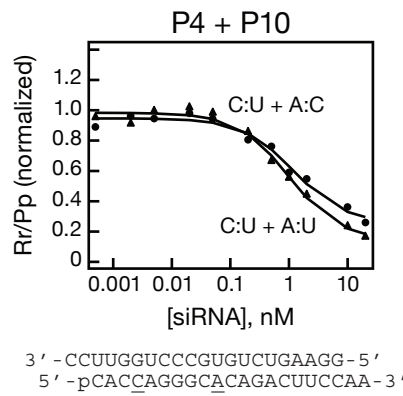
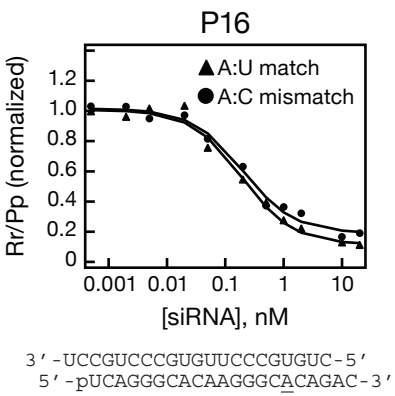
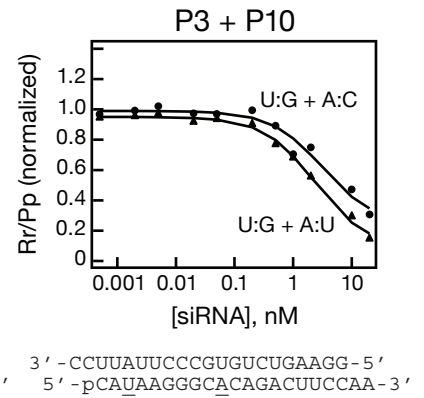
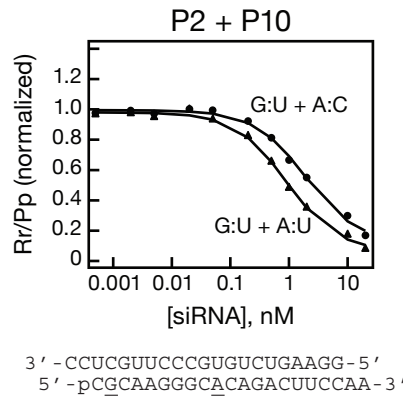




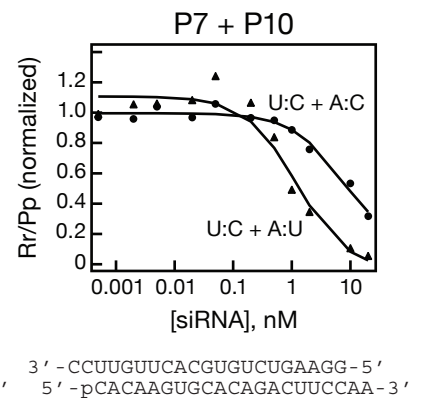
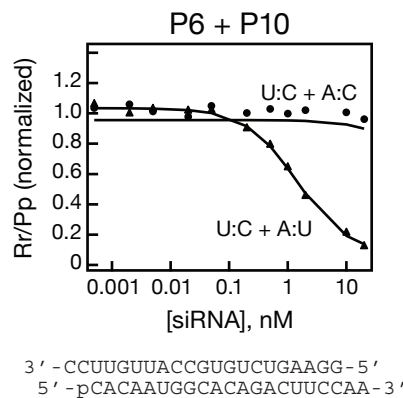
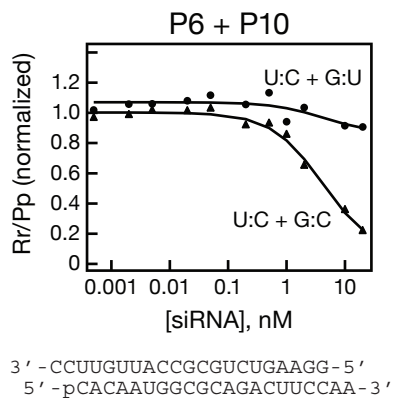
A



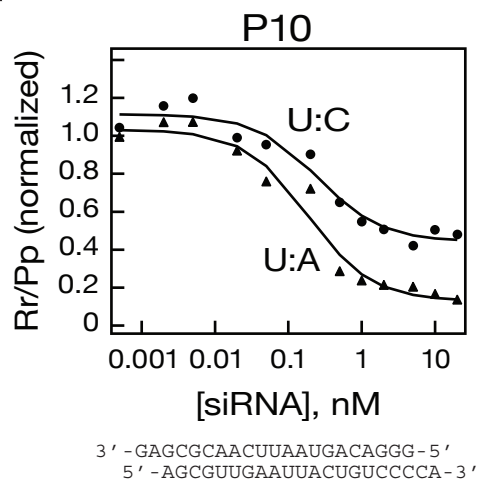
B



C



A



B

