A Human SNP in the *HTT* promoter alters NF-κB binding and is a Bidirectional Genetic Modifier of Huntington Disease

Kristina Bečanović[1,6*], Anne Nørremølle[2], Scott J. Neal[1], Chris Kay[1], Jennifer A. Collins[1], David Arenillas[1], Tobias Lilja[5], Giulia Gaudenzi[5], Shiana Manoharan[1], Crystal Doty[1], Jessalyn Beck[1], Nayana Lahiri[3], Elodie Portales-Casamar[1], Simon C. Warby[1], Colúm Connolly[1], Rebecca A.G. De Souza[1], REGISTRY Investigators of the European Huntington's Disease Network, Sarah J. Tabrizi[3], Ola Hermanson[5], Douglas R. Langbehn[4], Michael R. Hayden[1], Wyeth W. Wasserman[1] and Blair R. Leavitt[1*].

[1]Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, Department of Medical Genetics, University of British Columbia, Vancouver, BC, V5Z 4H4, Canada.
[2]Department of Cellular and Molecular Medicine, University of Copenhagen, DK-2200, Copenhagen, Denmark.
[3]UCL Institute of Neurology, University College London, Queen Square, London, WC1N 3BG, UK.
[4]Department of Psychiatry and Biostatistics, University of Iowa, Iowa City, 52242-1009, IA, USA.
[5]Department of Neuroscience, Karolinska Institutet, 171 76, Stockholm, Sweden.
[6]Department of Clinical Neuroscience, Karolinska Institutet, 171 76, Stockholm, Sweden.

* Corresponding Authors:
Kristina Bečanović Ph.D.
Department of Clinical Neuroscience
Karolinska Institutet, 171 76, Solna, Sweden
Email: Kristina.Becanovic@ki.se

Blair R. Leavitt MD.CM
Department of Medical Genetics
Centre for Molecular Medicine & Therapeutics
Child and Family Research Institute
University of British Columbia
Vancouver, BC, Canada
Email: bleavitt@cmmt.ubc.ca

ABSTRACT

Cis-regulatory variants that alter gene expression can modify disease risk, onset, and severity, but none have previously been identified in Huntington Disease (HD). Here we provide *in vivo* evidence in HD patients that *cis*-regulatory variants in the *HTT* promoter are bidirectional modifiers of HD age of onset. *HTT* promoter analysis identified an NF-κB binding site regulating *HTT* promoter transcriptional activity. A non-coding SNP, (rs13102260:G>A), within this binding site impaired NF-κB binding, reduced *HTT* transcriptional activity and HTT protein expression. Presence of the rs13102260 minor (A) variant on the HD disease allele was associated with delayed age of onset in familial cases, while presence of the rs13102260 (A) variant on the wild-type *HTT* allele was associated with earlier age of onset in HD patients in an extreme case based cohort. Our findings provide a novel mechanism linking allele-specific effects of rs13102260 on *HTT* expression to HD age of onset, and have significant implications for *HTT* silencing treatments in development.

Huntington Disease (HD) is an autosomal dominant neurodegenerative disease caused by a polyglutamine-encoding CAG repeat expansion in exon 1 of the huntingtin gene (*HTT*). While huntingtin is ubiquitously expressed in most tissues[1,2], the neuropathology of HD is characterized by relatively selective neuronal cell death in the striatum and cortex[3]. Wild-type huntingtin plays a crucial role in development of the nervous system and is protective against various forms of cytotoxicity, including neurotoxicity induced by mutant huntingtin[4,5]. The cytotoxicity of mutant huntingtin has been proposed to be the result of several mechanisms, including the disruption of axonal transport, mitochondrial dysfunction, imbalance in calcium homeostasis and excitotoxicity, altered proteolysis, impairment of the ubiquitin proteasome system and altered transcription[6]. The balance of expression between mutant and wild-type huntingtin may be an important modulator of pathogenesis and disease progression in HD.

In HD, the length of the CAG repeat in the expanded *HTT* disease allele inversely correlates with age of onset (AO). Although CAG repeat length accounts for a large proportion of the variation in AO[7], there is significant variability observed between expected and observed AO in HD patients[8]. HD patients with identical CAG repeat lengths, particularly those with CAG sizes between 40-44, can have AOs that differ by more than 20 years[9]. Increasingly, disease-associated non-coding single-nucleotide polymorphisms (SNPs) are being identified as potential modifiers of disease progression[10,11]. *Cis*-acting regulatory SNPs (rSNP) can alter the binding affinity of transcription factors to their binding sites, changing gene transcriptional activity and consequently modifying disease phenotype or progression. Other modifiers such as interacting genes and environmental factors likely also contribute to the variation in AO

observed in HD[12]. Both *cis*- and *trans*-acting factors that regulate *HTT* gene expression are likely to affect AO and disease progression in HD.

Genetic variation in regulatory sequences can affect the binding of transcription factors (TF) and alter the rate of transcription; such regulatory sequence variations could affect the expression of the *HTT* gene leading to functional consequences in HD. The expression levels of both wildtype and mutant huntingtin have been shown to modify the disease phenotype in many HD models[13-16]. The partially characterized murine *Htt* and human *HTT* promoters have a high GC content and contain a number of consensus binding sites for known TFs such as Sp1, AP2 and p53[17-20].

To determine whether genetic variation in the human *HTT* promoter influences the transcriptional rate of *HTT*, we established a panel of *HTT* promoter reporter constructs originating from HD patients We identified a NF-┤ B binding site containing a human SNP (rs13102260:G>A) that modulated the binding of NF-┤ B, resulting in reduced transcriptional activity of the *HTT* promoter in reporter gene assays. In order to investigate this SNP for disease-modifying effects, we assessed HD patient cohorts. We applied family-based designs and extreme phenotype sampling (EPS) using case extremes to increase efficiency in testing this candidate SNP for effects on HD age of onset[21-26]. As we show here, familial case and EPS based designs compared to random sampling, provided more powerful strategies for testing low frequency variant effects by enriching for genetic effects. In familial cases, we observed that the minor (A) sequence variant phased to the HD disease allele was associated with delayed age of onset as compared to the (G) sequence variant. In contrast, in our EPS cohort, we found that the (A) variant associated with an earlier AO when phased to the wild-type allele. Finally, we

showed that the genotype at rs13102260 regulates HTT protein levels correlating with the observed SNP effects on HD age of onset. In this study we provide *in vivo* evidence that *cis*-regulatory variants in the *HTT* promoter region can act as bidirectional modifiers influencing age of onset in HD with allele-specific effects, and that this effect was predicted by our *in vitro* data showing the direct effect of rs13102260 on NF-| B binding and huntingtin expression.

RESULTS

Cis-acting variants alter transcriptional activity of the human *HTT* promoter

Genomic regulatory sequences and binding motifs play key roles in transcriptional regulation. Genetic variants within TFBS might affect the binding of TFs and subsequent up- or down-regulation of transcription. Recent studies suggest that transcription rate data is more informative than steady-state mRNA abundance to use in gene regulation analysis[27]. We engineered a panel of constructs to study the transcriptional activity of the human *HTT* promoter using the luciferase reporter system. DNA samples from four representative HD patient haplotypes were selected (Supplemental Data 1). The human *HTT* promoter from the YAC128 transgenic mouse model of HD was also cloned and included in the panel. This transgenic mouse carries the entire human *HTT* promoter and gene with an expanded CAG repeat inserted into a yeast artificial chromosome[28]. Twelve distinct *HTT* promoter constructs were generated, including the eight alleles from the four HD patients, three pseudo-alleles with random PCR-induced sequence variants, and one from the YAC mouse. Basal transcriptional activity of the twelve constructs was initially measured using the luciferase reporter system in a human embryonic kidney cell line

(HEK293A). Eleven constructs displayed similar levels of basal transcriptional activity, while construct 5 (a pseudo-allele) exhibited a significant reduction in transcriptional activity compared to the other constructs ($t(10)=4.32$, $P<0.01$) (Fig. 1a). Similar experiments were performed in the ST14A cell line, a conditionally immortalized rat striatal cell line displaying similarities to striatal medium spiny neurons[29,30], with comparable results ($t(10)=5.83$, $P<0.001$) (Fig. 1b). The transcriptional activity of construct 5 was significantly reduced compared to the parental construct 4 (approximately 50%) in both the HEK293A and the ST14A cells. We next sequenced the 3.7 kb *HTT* promoter region for all twelve constructs (data not shown). Sequence comparison between constructs 4 and 5 identified ten sequence differences in the non-coding region of the huntingtin promoter (Fig. 1c).

Reduced transcriptional activity by the proximal HTT promoter region

We used a modified form of the RAVEN (Regulatory Analysis of Variation in ENhancers) bioinformatics software[31] (http://www.cisreg.ca/RAVEN) to identify putative TFBS affected by the identified sequence variants. Our results prompted an investigation of known human SNPs overlapping the identified binding sites. Four human SNPs that overlap with the putative TFBS in this region were identified: rs35207696, rs28441493, rs184840072 and rs13102260 (Fig. 1c). rs35207696, rs28441493, and rs184840072 are rare genetic variations with no individual genotype data and no frequency submission (http://www.ncbi.nlm.nih.gov). In contrast, the rs13102260 (G>A) occurs in the human population with a minor allele frequency MAF (A)=0.158/792 (1000 Genomes Project Phase 3); (MAF) of 0.042 for CEU (Caucasian European); MAF=0.050 for CHB+JPT (Han Chinese; Japanese) and MAF=0.441 for YRI (African Yoruban)

(http://www.ncbi.nlm.nih.gov). As the constructs 4 and 5 differed in several loci throughout the promoter region, a sequence-swapping procedure was used to identify the segment in the promoter of construct 5 responsible for the reduced reporter expression shown in Figure 1a and 1b. The 3.7 kb promoter was divided into three consecutive promoter fragments, A, B and C, with fragment C immediately proximal to the *HTT* translation initiation codon (Fig. 1c). Each promoter fragment was sequentially swapped between construct 5 and construct 4. Swapping the most distal A and B fragments did not alter the baseline transcriptional activity of construct 4 (Fig. 1d). However, when swapped from construct 5 to 4, fragment C significantly reduced the transcriptional activit y of construct 4 to the level of construct 5. In contrast, swapping fragment C from construct 4 into construct 5 conferred normal reporter expression, rescuing the reduced baseline transcriptional activity of construct 5 ($F(5,12)=11.9$, $P<0.001$, ANOVA with Bonferroni's multiple comparison test). We next assessed if fragment C required the distal A and B fragments to be present in order to have this modulatory effect on expression. Isolated fragment C from both constructs 4 and 5 showed similar levels of transcriptional activity as the full-length constructs they originated from ($F(3,8)=41.0$, $P<0.001$, ANOVA with Dunnett's multiple comparison test) (Fig. 1e). From these results we concluded that the proximal promoter region of fragment C (~1.1kb) contains *cis*-expression QTL (eQTL) that regulate gene expression, and that construct 5 contains genetic variation in this region that significantly reduces *HTT* expression.

Cis-regulatory SNP in NF-κB binding site reduces transcriptional activity

The alignment between constructs 4 and 5 revealed only three sequence variants in the identified eQTL in the 1.1 kb proximal *HTT* promoter region. *In silico* analysis further

predicted these variants to overlap with binding sites for CF1/USP, AML-1 and NF-┤ B (Fig. 1c). The predicted TFBS, the sequence variant position relative to the translation initiation codon start site, and the sequence variants for constructs 4 and 5 were, respectively: CF1/USP, -763 (T/C); AML-1, -392 (A/G); and NF-┤ B, -139 (G/T) (Fig. 1c). We next investigated the functional significance of altered binding of the identified TFs on transcriptional regulation of the *HTT* promoter. We performed site-directed mutagenesis to introduce the three identified construct 5 sequence variants into construct 4. The resulting promoter constructs revealed that changes at the predicted CF1/USP and the AML-1 binding sites did not affect the transcriptional activity in the reporter assay (Fig. 1f). In contrast, the construct 5 variant of the NF-┤ B binding motif significantly decreased the transcriptional activity of construct 4C to the levels observed for the construct 5 full-length and fragment C constructs ($F(4,23)=43.6$, $P<0.001$, ANOVA with Dunnett's multiple comparison test). The single base pair difference in the NF-┤ B binding sites of promoter constructs 4 and 5 was in the tenth and last nucleotide of the T FBS (Fig. 1g). These results suggested that the construct 5 variant of the NF-┤ B binding site caused the reduction in transcriptional activity.

We located the rs13102260 (G>A) at the first base position of the NFκB TFBS consensus motif from our database search on SNPs overlapping the identified putative TFBS in the *HTT* promoter region (Fig. 1g). We thus identified a human SNP that was present in the same NF-┤ B binding site, but that was distinct from the nucleotide variation that we had originally investigated *in vitro*. Site-directed mutagenesis was performed to introduce the rs13102260 variant into construct 4C in order to study its effect on transcriptional activity. The rs13102260 (A) variant caused an approximate 50%

reduction in transcriptional activity similar to constructs 5 and 5C ($F(5,14)=26.2$, $P<0.001$, ANOVA with Dunnett's multiple comparison test) (Fig. 1h). Thus sequence changes at both the first and last position of the NF-┤ B binding site resulted in significant reduction in transcriptional activity of the *HTT* promoter.

Activated NF-┤ B binds to the *HTT* promoter *in vitro* and *in vivo*

We next performed a series of biochemical experiments to determine the binding characteristics of the identified TFs and to functionally validate NF-┤ B binding to the *HTT* promoter *in vitro* and *in vivo*. We performed chromatin immunoprecipitation (ChIP) analysis to study the putative NF-┤ B occupancy of the huntingtin promoter. Enrichment of NF-┤ B binding to the predicted TFBS was assessed using quantitative real-time PCR (qPCR). To investigate whether NF-┤ B bound the site *in vivo*, we first performed ChIP using different brain regions of naïve mice (Fig. 2a-c). The mouse *Htt* promoter contains three putative NF-┤ B TFBS within 1000 bp upstream of the translation initiation codon (Supplementary Fig.1a). There was a clear NF-┤ B enrichment in all brain regions and at all analyzed TFBS regions comparable to what was observed for the IL6 promoter (positive control for NF-┤ B binding) (striatum $F(7,32)=1.75$, $P=0.134$; pre-frontal cortex $F(7,32)=2.44$, $P<0.05$; cerebellum $F(7,32)=7.22$, $P<0.001$, one-way ANOVA). Interestingly, there was a significantly higher level of NF-┤ B recruitment in striatum compared to pre-frontal cortex and cerebellum at TFBS2 (Tissue $F(2,96)=5.53$, $P<0.01$; TFBS $F(7,96)=3.15$, $P<0.01$; interaction $F(14,96)=1.17$, $P=0.311$, two-way ANOVA) (Fig. 2d). The similar level of enrichment at each TFBS is likely due to limited ChIP resolution (200-1000bp). To study the NF-┤ B occupancy on the *HTT* promoter in a

human context, ChIP analysis was performed in lymphoblastoid cell lines (LCL) derived from HD patients (Supplementary Fig. 1b). There was an increased enrichment of NF-┤ B at the single predicted NF-┤ B TFBS (comprising rs13102260) in human LCLs ($F(3,32)=4.28$, $P<0.05$, one-way ANOVA with Dunnett's multiple comparison test) compared to a region upstream in the *HTT* promoter region that we used as a control containing two putative NF-┤ B TFBS (-2065; -2011bp from translation start site) (Fig. 2e). Rat striatal ST14A cells, which is the system employed in the reporter assays, were then stimulated with TNFα to activate the NF-┤ B pathway. The rat *Htt* promoter contains four putative NF-┤ B TFBS within 1000 bp upstream of the translation initiation codon (Supplementary Fig.1c). We assessed the binding of NF-┤ B to the distinct TFBS using three different primer sets. We observed an enrichment of NF-┤ B for all three regions comparable to the IL6 promoter ($F(7,16)=5.09$, $P<0.01$, one-way ANOVA with Bonferroni's multiple comparison test) (Fig. 2f). These results thus demonstrate an enrichment of NF-┤ B occupancy at the predicted NF-┤ B TFBS proximal to the TSS both *in vitro* in immortalized rat striatal ST14A cells and in LCLs derived from HD patients, as well as *in vivo* in mouse striatum, pre-frontal cortex and cerebellum. Notably, we observed a significantly higher level of NF-┤ B recruitment in striatum compared to pre-frontal cortex and cerebellum.

rSNP alters NF-┤ B binding to the *HTT* promoter

As shown in figure 1, sequence variants at both the first and the last position of the NF-┤ B binding site resulted in significant reduction in transcriptional activity of the *HTT* promoter. We used electrophoretic mobility shift assays (EMSA) to more precisely

investigate the interactions between NF-┤ B and the binding site harboring the rSNP. NF-┤ B is predominantly cytoplasmic under normal conditions, but migrates to the nucleus upon activation by cytokines such as TNFα. We therefore performed TNFα stimulations to activate the NF-┤ B pathway in ST14A cells prior to nuclear extract preparation. In the EMSA, NF-┤ B bound more strongly to the allele 4 compared to the allele 5 oligonucleotide in a concentration-dependent manner (Fig. 3a). The shifted band was abolished upon addition of unlabeled competitor oligonucleotide. The oligonucleotides (allele 5 oligonucleotide in particular) containing the NF-┤ B binding site were prone to create secondary structures such as self-dimers or hairpin structures. The NF-┤ B protein, however, also exhibited a high affinity to these secondary structures as well as to the duplexed oligonucleotide (data not shown). In the presence of recombinant human NF-┤ B p50 protein, a single shifted band was observed that increased in intensity in a concentration-dependent manner (Fig. 3b). Addition of 0.9 gel shift units (gsu) produced a single band that was stronger for the allele 4 oligonucleotide than was observed for the allele 5 oligonucleotide. The allele 5 band was out-competed more efficiently with unlabeled competitor oligonucleotide, indicating lower binding affinity to the allele 5 compared to the allele 4 oligonucleotide. EMSA confirmed that NF-┤ B binding to the oligonucleotide containing the rs13102260 (A) variant was fully abolished, as assessed with both ST14A nuclear extract (Fig. 3c) and recombinant NF-┤ B p50 protein (Fig. 3d). The binding of CF1/USP and AML-1 to the allele 4 and allele 5 target sequences was also assessed following incubation with ST14A nuclear extract. None of these showed decreased binding to the allele 5 target sequence (Supplementary Fig. 2). The p50/p65 heterodimer is the most abundant of the NF-┤ B dimers, displaying a very potent gene

regulatory function. Different DNA targets determine the final conformation of the heterodimer[32,33]. The p50/p65 dimer complex specifically recognizes the 5'-GGGRNYYYCC-3' (R=unspecified purine; N=any nucleotide; Y= unspecified pyrimidine). The p50 subunit strictly recognizes 5'-GGGRN-3', while the p65 subunit recognizes 5'-YYCC-3', but at less stringency. These results suggest that the p50/p65 heterodimer still binds to both the construct 4 and 5 variant at the tenth base position in the NF-┤ B TFBS, but to a lesser extent due to reduced binding of the p65 subunit. In contrast, the rs13102260 (A) variant at the first position in the NF-┤ B TFBS affects the binding of the p50/p65 heterodimer to a higher degree by abolishing the p50 binding to the *HTT* promoter. These findings demonstrate that sequence variants at the first as well as last base pair of the NF-┤ B binding motif in the *HTT* promoter alter NF-┤ B binding, subsequently reducing transcriptional activity of the *HTT* gene.

Targeting of NF-┤ B modulates *HTT* expression

We next investigated whether the NF-┤ B-mediated transcriptional activity on the *HTT* promoter was dynamic and therefore a potential therapeutic target. We first assessed whether siRNA knockdown of NF-┤ B influenced *HTT* gene expression. This experiment showed that down-regulation of NFKB1 (specific for p105, precursor to the p50 subunit) and p65 mRNA levels, both of which encode critical components of the potent heterodimer p50/p65 complex, decreased the levels of *HTT* expression ($t(6)=3.52$, $P<0.01$) (Fig. 4a). The knockdown of NF-┤ B was also associated with decreased p53 mRNA to levels comparable with the decrease in *HTT* expression, p53 being an established NF-┤ B target ($t(6)=3.46$, $P<0.01$, Supplementary Fig. 3a). We next stimulated

ST14A cells with TNFα to assess if the treatment would increase the transcriptional activity of the *HTT* promoter constructs. Upon TNFα-stimulation, construct 4 exhibited increased transcriptional activity ($F(3,8)=29.7$, $P<0.001$, one-way ANOVA with Dunnett's test), in contrast to construct 5 that exhibited no change in expression compared to the untreated control (Fig. 4b). Caffeic acid phenethyl ester (CAPE), an active component of propolis from honeybee hives, is known for its anti-mitogenic, -carcinogenic, -inflammatory and immuno-modulatory properties[34]. CAPE blocks TNFα-induced activation of NF-┤B in a dose and time-dependent manner by preventing the translocation of the p65 subunit to the nucleus and by selectively inhibiting NF-┤B binding to DNA[35]. Stimulation of ST14A cells expressing construct 4 with TNFα in combination with CAPE treatment resulted in reduced transcriptional activity of the promoter in a CAPE dose-dependent manner ($F(4,7)=6.24$, $P<0.05$, one-way ANOVA with Dunnett's test) (Fig. 4c). We next assessed the effects of CAPE treatment on the basal transcriptional activity of the *HTT* promoter constructs. Neither constructs 4 nor 5 displayed changes in transcriptional activity levels when the ST14A cells were treated with CAPE in the absence of a NF-┤B-pathway activating treatment (Supplementary Fig. 3b). Western blot analysis was performed to confirm a TNFα-induced increase in nuclear levels of NF-┤B protein in our experimental setting. TNFα-stimulated ST14A cell nuclear extract expressed all NF-┤B subunits, with the most dramatic increase in the p50 subunit compared to untreated cells (Fig. 4d). The cytoplasmic fraction showed strong bands for the precursor subunits p100/p105, in addition to bands for the p50 subunit. There were, however, no bands for the p65, RelB and c-Rel subunits in the cytoplasmic fraction. Western blot analysis revealed that CAPE treatment did not affect the nuclear

p65 protein levels in the ST14A cells (Fig. 4d). Our results thus suggest that CAPE inhibits NF-┤B nuclear activity, perhaps through prevention of p65 binding to the *HTT* promoter. In summary, our results showed that siRNA knockdown of NF-┤B, TNFα and CAPE stimulations modulated the transcriptional activity of the *HTT* promoter.

rs13102260 (A) variant on the HD disease allele is protective and associated with delayed age of onset

Based on our *in vitro* findings we next studied the disease-modulatory effects of the rs13102260 (A) variant in HD patients. We previously showed that a specific HD haplotype was associated with later AO in Danish HD families[36]. Here we genotyped the familial cases to determine if the rs13102260 genotype was associated with delayed age of motor onset. We performed direct sequencing of HD subjects (n=98) originating from 36 Danish HD families representing seven different haplotypes including the one associated with later AO. Direct sequencing revealed that all HD subjects with the haplotype that we previously showed to associate with later AO (B-haplotype), exclusively carried the rs13102260 (A) variant. Phasing of the (G) and (A) sequence variants was done in each HD family by segregation of affected and/or unaffected genotypes within the pedigree. We found that the (A) variant phased to the HD disease allele in all B-haplotype families (n=8). The other HD haplotypes that were not associated with later AO carried the (G) variant. We next performed association testing to further evaluate and quantify the effect of the rs13102260 (G/A) on AO in these HD patients. We restricted our analysis to HD subjects with a CAG-length of 41-55, including 35 (G/A) heterozygotes with the (A) phased to the HD disease allele, 41 patients that were (G/G) homozygotes and six (A/A) homozygotes at rs13102260 (n=82)

(Supplementary Table 1). Analysis was performed using weighted least square models predicting observed ages of onset as a function of the (G/A) rs13102260 genotype, and expected age of onset based on the CAG repeat length as estimated by Langbehn *et al*[8]. Presence of the rs13102260 (A) variant on the HD disease allele had a clear association with age of HD onset, after first adjusting for expected age of onset based on CAG repeat length (Fig. 5a). The resulting mean delay of age of onset was 9.3 years for (A/G) heterozygotes at the mean expected age of onset in the data (40.1 years) compared to (G/G) homozygotes (t=5.08, 27.1 df, P<0.0001). The overall rSNP effect (main effect plus interaction) was statistically significant at P<0.0001 (F=14.40, df=2, 34.7). There was evidence for interaction between rs13102260 and the CAG-based expected age of onset (F=5.80, df= 1, 73.9, P=0.019) with the protective effect being stronger at lower CAG lengths (Fig. 5a, Supplementary Fig. 6). There was however no evidence of association in homozygous rs13102260 (A/A) individuals (adjusted mean difference: 5.56 years, t=1.90, df=53.4, P=0.063) (n=6). Hence, our results showed that the rs13102260 (A) variant had a protective role and associated with delayed AO when phased to the HD disease allele.

rs13102260 (A) variant on the wild-type allele associated with earlier HD age of onset

We next genotyped an additional cohort of HD patients (UBC HD Biobank) in an effort to validate the observed disease-modulatory effects of the rs13102260 (A) variant on HD age of onset. We used a family-based design to collect the HD subjects. The patients included in this cohort were assessed individually for age of motor onset, in addition to being mapped into family pedigrees. We further applied an extreme phenotype sampling

approach in which we selected the most extreme subject from each pedigree *i.e.* on the basis of age of onset where the ratio between observed age of onset and expected (based on CAG-length)[8,37] was used to identify early, mean or late age of onset HD subjects (see material and methods). This selection of case extremes, similarly to familial cases, enriches for alleles that contribute to disease, allowing detection of genotype associations in modest sample sizes. A total number of 459 HD subjects were genotyped for the rs13102260, of which only 28 out of 459 (approx. 6%, corresponding to MAF in Caucasians) were (G/A) heterozygotes, while the remainder where (G/G) homozygotes. We then genotyped relatives of the probands that were identified as (G/A) heterozygotes to phase the (A) variant to the wild-type or the HD disease allele. The rs13102260 (A) variant was almost exclusively 1) on the wild-type allele and thus 2) inherited from the unaffected parent (except for two subjects with the (A) variant phased to the HD disease allele) (Supplementary Table 2). As for the Danish cohort, we restricted our analysis to HD subjects with a CAG-length of 41-55, including 22 (G/A) heterozygotes with the (A) variant phased to the wild-type allele and 391 patients that were (G/G) homozygotes (n=413) (Supplementary Table 3). Analysis was performed using similar weighted least square models as for the Danish cohort. Presence of the rs13102260 (A) variant on the wild-type allele associated with age of HD onset, where subjects carrying the rs13102260 (A) variant displayed on average an age of onset 3.88 years earlier than those with the rs13102260 (G) variant (t=-2.57, df=367, P=0.010) (Fig. 5a). There was no evidence for interaction between the rs13102260 and the CAG-based expected age of onset. In addition, given that the UBC data were collected on the basis of age of onset, with the SNP genotype as the unknown variable being predicted, we also used a logistic

regression model to predict SNP genotype as the outcome. The difference between the observed and expected age of onset (CAG-based) was the main predictor of allele frequency. Results from this logistic regression were consistent with the results from the above described analysis. The adjustment for expected age of onset produced estimates of the odds ratio. The odds of having the (A) variant decreased by a factor of 0.929 per year increase in observed age of onset (F=5.75, df=1, P=0.017, two-sided Wald test, 95% CI= 0.875-0.987). In other words, the (A) variant phased to the wild-type allele was associated with an earlier age of onset regardless of CAG-repeat length. Our result thus suggests that the rs13102260 (A) variant is a genetic risk variant that accelerates AO in HD patients when phased to the wild-type *HTT* allele.

It should be noted that we do not have access to additional HD patient data of comparable sample size that are also enhanced by family-based designs or oversampling extreme-phenotype cases. We did attempt a third association study of a random population sample obtained from the EHDN Registry (n=26 (G/A); n=418 (G/G)) for which we did not observe a significant association between age of onset and the presence of the rs13102260 (A) variant on the wild-type allele (t=0.58, 324 df, P=0.56) (Supplementary Fig. 4). To our knowledge, no other HD patient population except the Danish cohort has been described to display a high frequency of the rs13102260 (A) variant phased to the HD disease allele. We observed an enrichment of (A) variant carriers with extremely early AO in the UBC population: 41% (9/22) of the HD subjects that displayed extremely early AO ($<15^{th}$ percentile) also carried the (A) variant (Fig. 5c), whereas 15% (4/26) of the subjects displaying an extremely early AO carried the (A) in the random EHDN Registry population of a similar size (Supplementary Fig. 4).

Genotyping familial cases in the Danish cohort also reflected enrichment effects with 26% (9/35) of the (A) variant carriers on the HD disease allele displaying extremely late age of onset (>85th percentile) compared to 9% (40/432) in the (G/G) carriers (Fig. 5c).

Estimation of rs13102260 (A) variant disease modifier effects on HD disease allele

Association analysis of the UBC cohort showed that HD subjects with the rs13102260 (A) variant displayed on average 3.9 years earlier age of onset compared to HD subjects with the (G) variant phased to the wild-type allele (Fig. 5b). In contrast to the UBC cohort, the relationship between CAG-repeat size and expected age of onset was non-linear for the Danish cohort as described (P=0.019) (Supplementary Fig. 6). This effect increased by 0.47 years per year of increased expected onset age (t=2.41, 73.9 df, P<0.05). The estimated disease-modifying effect of the (A) variant phased to the HD disease allele was greater in HD subjects with shorter CAG repeat lengths (Table 1). For example, (G/A) heterozygotes with CAG repeat lengths of 41 presented age of onset on average 17.3 years later than expected (extrapolation with no data for (G/A) heterozygotes), while (G/A) heterozygotes with CAG repeat lengths of 55, presented age of onset in average 2.7 years later. (G/A) heterozygotes with CAG repeat lengths of 43, which represent the shortest CAG-length with relevant data, presented a delayed AO with on average 13.1 years compared to (G/G) homozygotes. Here, to further show the effect of rs13102260 on age of onset, we calculated the ratio AO (=observed AO/expected AO)[8]. Compared to the (G/G) homozygotes, HD subjects with the (A) variant phased to the wild-type allele showed lower ratio AO values, while HD subjects with the (A) variant phased to the HD disease allele showed higher ratio AO values (Fig. 5d).

rs13102260 (A) variant correlates with distinct haplotype pattern

We next performed sequence analysis of the 22 (G/A) carriers analyzed in the UBC cohort. Interestingly, sequencing confirmed two distinct haplotype patterns correlating with the rs13102260 (A) variant including the 6bp repeat VNTR in the *HTT* gene promoter and the polymorphic CCG region in exon 1, which have been previously reported[36,38-40]. Of all the HD subjects with the (A) variant on the wild-type allele, 55% (12/22) carried the two-repeat allele at the 6bp VNTR loci in conjunction with $(CCG)_7(CCT)_3$ at the CCG polymorphic region (alternately referred to as CCG8) (Supplementary Table 4) (Supplementary Fig. 5). Forty-five percent (10/22) carried the one-repeat allele of the 6bp VNTR locus in conjunction with $(CCG)_7(CCT)_2$ at the CCG polymorphic region (alternatively referred to as CCG7). Previous studies showed that the copy number at the 6bp VNTR loci in itself did not affect transcriptional activity[17]**.** Remarkably, the haplotype carrying the rs13102260 (A) variant, the two-repeat alleles at the 6bp VNTR loci and CCG8 on the wild-type *HTT* allele that was represented in the (G/A) heterozygotes in the UBC cohort that associated with earlier AO, corresponded to the haplotype on HD disease alleles in the Danish HD families that we previously reported to be associated with delayed AO[36].

Genotype at rs13102260 regulates allele-specific expression of HTT protein

Finally, we assessed whether the genetic variation at rs13102260 modulates protein levels in HD patient fibroblast lines. We compared HD patient fibroblast lines with different genotypes at rs13102260: homozygous (G/G), heterozygous with (A) variant phased to the wild-type allele or heterozygous with (A) variant phased to the HD disease allele, respectively. The (A) variant significantly down-regulated HTT protein levels when

phased to the wild-type allele ($t(3)=3.27$, $P<0.05$) (Fig. 6a). The HTT levels were reduced by approximately 50%, which corresponds to the reduction observed in transcriptional activity (Fig. 1h). As expected, the (A) variant also significantly down-regulated mutant HTT (mHTT) protein levels when phased to the HD disease allele (~25% decrease, $t(3)=2.46$, $P<0.05$) (Fig. 6a). We observed a general reduction of mHTT protein levels compared to the wild-type protein regardless of genotype. This allele-specific expression difference might be explained by intrinsic tissue-specific effects, epigenetic modifications or HD haplotypes that might contain *cis*-regulatory elements influencing gene regulation. Thus, we showed that the (A) variant is a modifier of protein levels; the (A) variant phased to the wild-type allele significantly reduced wild-type HTT protein levels, while the (A) variant phased to the HD disease allele reduced the levels of mHTT protein. These results revealed functional consequences of this *cis*-acting regulatory variant that altered NF-┤ B binding, reduced HTT protein levels and associated with modulation of age of onset in HD patients (Fig. 6b).

DISCUSSION

Genetic disease modifiers cause single gene diseases such as HD to display features of complex traits, by influencing the disease expressivity[41,42]. We here identify a *cis*-acting regulatory variant in the *HTT* promoter that acts in an allele-dependent manner as a bidirectional genetic modifier in HD. We show that the identified rs13102260 (G/A) SNP decreases NF-┤ B binding to the *HTT* promoter resulting in reduced HTT expression and exerting allele-specific modifier effects on age of onset in HD patients.

To put our results into clinical context, patients with the (A) variant phased to the HD disease allele develop motor symptoms on average 10 years later compared to HD

patients with the (G) sequence variant. Interestingly, the estimated disease-modifying effect was greater in HD subjects with shorter CAG repeat lengths and declined with increased CAG repeat length (Table 1). In contrast, we further show in an independent cohort that patients with the (A) variant on the wild-type allele are estimated to develop HD motor symptoms almost 4 years earlier than patients with the (G) variant. SNP analysis and genotype association tests thus demonstrated the disease-modulatory effects of rs13102260 (G/A) on AO in HD. Biochemical analysis further validated the potency of this *cis*-acting genetic variant by measuring its allelic-specific effects on transcriptional activity, binding- and protein expression.

Several potential *trans*-acting disease modifiers which influence AO in HD have previously been reported[43-47]. Although, none of the reported findings has yet revealed the mechanism of action by which it contributes to disease pathogenesis, identification of genetic modifiers in HD is important for two reasons: 1) these genes and related pathways are excellent potential therapeutic targets in HD; and 2) these may provide improved prognostic information for HD gene carriers.

Genome-wide studies have demonstrated that differential allelic gene expression is common *i.e.* allelic imbalance[48]. *Cis*-regulatory variants altering gene expression are an important potential source of phenotypic difference thought to play an important role in the pathogenesis of many complex diseases, but so far none have been identified in HD. Recently, a genome-wide study focusing on NF-┤ B (p65) binding sites showed that genomic structural variants such as SNPs frequently caused differences in gene expression due to altered binding of NF-┤ B[49].

Although NF-┤ B is nearly ubiquitously expressed, its role in neurons of the central nervous system is controversial[50] (see Supplemental Discussion). We showed that NF-┤ B regulation of the *HTT* promoter was dynamic both using siRNA knockdown of NF-┤ B and with cytokine stimulations and inhibitors. Interestingly, we also showed an increased NF-┤ B activity in striatum in naïve mouse brain, which is the brain region primarily affected in HD. Our overall findings demonstrate that NF-┤ B directly binds to and increases the transcriptional activity of the *HTT* promoter; providing the potential for a toxic feed-back loop which could result in a detrimental increase in expression of mutant HTT in neurons in the striatum. Given the selective effects of HD pathogenesis in the striatum, our finding that NF-┤ B activity is significantly increased in striatum in naïve mouse brain suggest that the interplay between NF-┤ B and HTT merits additional investigation. We propose NF┤ B to be further investigated as a therapeutic target in HD.

Expression levels of mutant HTT contribute to neuropathology in HD, most probably in combination with an increased sensitivity in specific brain regions[13]. The expression levels of mutant Htt modulate both the onset and progression of the HD phenotype in the YAC128 mouse model of HD. Increased levels of mutant Htt are associated with increased sensitivity to excitotoxicity, earlier AO and more rapid disease progression[14]. Importantly, in this mouse model of HD, decreasing the levels of wild-type Htt increases the cellular toxicity of mutant Htt[15,16], and over-expression of wild-type Htt ameliorates striatal neuronal atrophy[51].

Cis- and *trans* acting factors as well as different environmental exposures alter the balance of gene expression, and ultimately affect disease expressivity in HD patients[41,42]. Recent studies suggest allelic imbalance with mutant *HTT* mRNA more abundant

compared to the wild-type mRNA in post-mortem HD striatum and cortex, while there was no difference in the cerebellum[52]. Our human data suggest that reduced levels of mHTT delay the onset of HD, providing support for future *HTT* gene silencing therapeutic approaches. Our data also suggest that reduced levels of wild-type HTT may accelerate HD onset, although the effect of the rs13102260 genotype appears to be smaller. The balance between wild-type and mutant HTT could be thought of as a "yin-yang" relationship. We here demonstrated that the *cis*-acting rs13102260 sequence variant potently affects the regulation of the *HTT* promoter by altering the NF-κB binding site, suggesting that the sequence (A) variant leads to increased allelic imbalance. Calculations on the estimated rSNP effect further suggest that the *cis*- regulatory (A) variant exerts a much greater disease-modifying effect on AO when phased to the HD disease allele compared to the wild-type *HTT* allele. Our *in vivo* results in HD patients thus suggest that reduction of mHTT is a more potent disease modifier of AO than reduction of wild-type HTT. HD subjects with CAG repeat lengths of 43 presented AO on average 13.1 years later than expected (Table 1). This has implications not only for our understanding of the interplay between wild-type and mutant HTT, but more importantly for the gene silencing strategies applied in HD. Our results thus highlight the importance of continued refinement of our understanding of allelic imbalance in HD patients.

HD is a single gene disorder thought to be mediated by a toxic gain-of-function, and silencing of *HTT* expression is an attractive therapeutic approach. Non-allele-specific silencing of both the wild-type and the mutant HTT in HD fibroblast lines resulted in increased caspase-3-like activity, while knockdown of only the mutant HTT did not[53].

Studies performed in mice have suggested that non-allele specific silencing alleviates HD symptoms[54-56], while other studies have provided support for allele-specific silencing in HD[57-59]. Our results suggest a disadvantage of lower wild-type HTT levels, and as most *HTT* silencing studies have been relatively short-term and performed in mouse, an issue of potential concern is whether non-allele specific silencing may have undesirable effects by decreasing levels of wild-type HTT. Neuroprotective disease-modifying treatments will likely start in young adulthood and continue throughout life. Our study nevertheless provides the first clinical evidence that *HTT* gene silencing therapies for HD will be efficacious. If these therapies are non allele-specific, we predict that they will have a net beneficial effect, but that allele-specific approaches targeting only mHTT may have greater therapeutic efficacy.

Identification of disease-modifying targets and/or mechanisms would be an invaluable and long awaited addition for drug development in HD as well as other neurodegenerative diseases. In this study, we have identified NF-┤ B as a trans-acting factor that regulates *HTT* gene expression and rs13102260 to act as a bidirectional disease modifier. We thus refine the notion of allele-specific relationships between *cis*-acting regulatory elements, the transcriptional rate of *HTT* expression, and age of onset in HD subjects. In conclusion, this study has implications for therapeutic strategies aimed at silencing of the *HTT* gene in HD patients. Genotyping of rs13102260 may further provide prognostic information of significance for a subset of HD gene carriers. Continued identification of *cis*- and *trans* regulatory elements will provide insights into their impact on disease expressivity and identify novel targets for disease-modifying therapeutic interventions in HD.

METHODS

Promoter constructs used in reporter assays

DNA samples from four HD patients from the Huntington Disease BioBank at the University of British Columbia (UBC HD BioBank) were selected that represent a variety of different haplotypes observed in HD patients (Supplementary information). All of these haplotypes carried the rs13102260 (G) sequence variant. The HD promoter fragments were amplified together with the disease causal polymorphic region of exon 1 to maintain the allelic linkage between the two elements. The PCR products were cloned to isolate individual clones of each allele in order to abrogate the possibility of monoallelic bias in sequencing. Furthermore, both strands of multiple clones per patient were sequenced using overlapping sequence reads to rule out any artefacts from amplification. Sequence analysis performed prior to cloning into the reporter constructs identified additional PCR-induced single base pair changes generated during the amplification process. Three pseudo-alleles displaying PCR-induced sequence changes were included in our panel to assess their potential effects on transcriptional activity (constructs 3, 5 and 8 were derived from parent constructs 2, 4 and 9, respectively). Human HD patient DNA was PCR-amplified using the following primers: forward 5' CTC AGA GAC ACC ATG CCAGA 3'; reverse 5'AGC CCT CTT CCC TCT CAG AC 3'. The PCR-amplified DNA was sub-cloned using the TOPO® XL PCR cloning kit (Invitrogen). PCR amplification of the CAG tract was used to determine if the clones originated from the wild-type or the CAG expanded allele. TOPO PCR vector plasmid was digested with *NcoI* (NEB) and fragments were purified before ligating DNA into the pGL3-basic promoterless luciferase reporter gene vector (Promega). Cloned full-length *HTT* promoter constructs were 3742 bp. The following restriction enzymes were used for

cloning: fragment A (HindIII-NdeI); fragment B (NdeI-PstI); fragment C (PstI-BstBI). Plasmid preparations were carried out using the endotoxin-free Maxi kit (Qiagen).

Reporter assays

HEK293A and immortalized striatal ST14A cells were used for the reporter assays. Cells were seeded onto 96-well white plates corresponding to ~80% confluency. Trans-LT1 transfection reagent (Mirus Inc.) was used to co-transfect the *HTT* promoter reporter constructs, the renilla vector (internal reference control) and pGL3 vector to a total DNA amount of 200 ng per well 24 hours post seeding of cells. The results are presented as experimental sample ratio where firefly luciferase activity is normalized to *Renilla* luciferase activity as expressed in relative light units (RLU). Negative control wells constituted empty pGL3 vector and renilla plasmid. Constructs and the constructs with changed TFBS were assayed in three and six wells, respectively. Cells were treated with the luciferase substrate reagents 24 hours post transfection (Dual-luciferase assay, Promega). The results are presented as experimental sample ratio where firefly luciferase activity is normalized to *Renilla* luciferase activity and is expressed in RLU. TNF⟨ was added to the cells six hours after transfection to a final concentration of 20 ng/ml where indicated. CAPE (TOCRIS bioscience) was diluted to different concentrations: 2.5ng/∝l, 10ng/∝l, and 20ng/∝l and fixed volumes (1∝l) were added two hours prior to stimulation with TNF⟨ (eBioscience).

***In silico* analysis**

The cloned DNA sequences were compared to the human build NCBI36, human hg18, March 2006 (http://www.ncbi.nlm.nih.gov). Identified sequence variants in the cloned constructs affecting the predicted TF binding sites were analyzed implementing a

modified version of the RAVEN software analysis (http://www.cisreg.ca/RAVEN). The construct sequences were scanned using position weight matrices obtained from the JASPAR database of high-quality transcription factor binding profiles (http://jaspar.genereg.net/)[60]. The binding site scores for each transcription factor in positions overlapping sequence variants were then compared between the two sequences. The parameters were set to high stringency, requiring binding site profile scores exceeding 80% of maximum on at least one of the two constructs. The highest scoring predicted TFBS overlapping a sequence variant position was compared to the highest scoring predicted TFBS overlapping the same sequence variant position on the other allele. TFBS affected by sequence variants were thus defined as any TFBS where the relative score of the TFBS profile matrix was at least 80% on either construct 4 or 5 and for which the absolute score difference between the binding sites on the two alleles exceeded 1.5.

Electrophoretic mobility shift assays (EMSA)

EMSA experiments were performed on nuclear extracts from ST14A cells and commercially manufactured biotinylated oligos (IDT). Nuclear extracts were prepared according to manufacturer's protocol (NE-PER® Nuclear and Cytoplasmic Extraction reagents, Pierce Thermofisher). Commercially available recombinant p50 was used (Promega). EMSAs were performed using double-stranded oligonucleotides spanning each putative site. ~30 bp oligonucleotides were designed constituting the respective binding sites for CF1/USP, AML-1 and NF-┤ B as described in Supplementary information. The forward oligo was biotinylated and duplexed to the unlabeled reverse oligo. EMSA reactions were performed according to manufacturer's protocol

(Lightshift® Chemiluminescent EMSA kit, Pierce Thermofisher). ST14A cells were stimulated with TNF‹ (eBioscience) (80 ng/ml for 24 hours) for the NF-┤ B gelshift assay.

Chromatin immunoprecipitation

ST14A cells were stimulated with TNFα (80 ng/ml) for 24 hours prior to fixation. Experiments were performed according to manufacturer's recommendations (Upstate, Millipore). Adult mice (NMR1 strain, 9 months of age; total n=5; male=3, female=2) were sacrificed and tissue from striatum, pre-frontal cortex and cerebellum was dissected out. Tissue was collected in PBS with protease inhibitors (Diagenode). Tissue samples were disrupted with syringe and needle and cells were fixed in 1% formaldehyde in PBS for 10 min before addition of glycine to a final concentration of 125 mM. Cells were pelleted by centrifugation at 500g for 5 min and washed twice in PBS. For human lymphoblastoid cells (LCLs); cells were harvested by centrifugation at 100g for 8 min and washed in PBS with protease inhibitors before fixation in 1% formaldehyde for 10 min followed by addition of glycine to a final concentration of 125 mM. LCLs were pelleted by centrifugation and washed twice with PBS. Nuclear extracts were prepared from LCL pellets and brain tissue samples following the instructions using the "high Cell# ChIP kit" (Diagenode). Chromatin shearing by sonication was performed using a Bioruptor UCD 200 (Diagenode). Shearing efficiency and chromatin concentration was evaluated after reverse crosslinking and DNA isolation. NFκB antibody (SC-114X, Santa Cruz) and control IgG antibody from the High Cell# ChIP kit (Diagenode) was used in all ChIP experiments. Results were normalized to 1% input. DNA was analyzed by qPCR on the 7500 Fast Real-Time PCR System (Applied Biosystems). Analysis was performed

using the 7500 Software V2.0.1 (Applied Biosystems). Data was normalized to input DNA. Primers are described in Supplementary information.

Site-directed mutagenesis

PCR was performed using HiFi Taq (Invitrogen) with the construct 4 C plasmid as template. Primers used are described in Supplementary information. *DpnI* was used to digest the methylated template leaving only the unmethylated PCR product. Sequence analysis was performed to select for the clones with the correct inserted mutation. Mutagenized clones were cloned into the pGL3 plasmid and transformed into electrocompetent DH5α.

Western blot analysis

Western blotting was performed on protein samples extracted from ST14A cells according to manufacturer's protocol (Thermofisher). ST14A cells were treated with either TNFα (eBioscience) (20 ng/ml) for 6 hours, or in combination with CAPE (TOCRIS bioscience) (175 ng/ml) that was added 1.5 hours prior to addition of TNFα. 19µg of total protein was boiled for 5 minutes before being separated on 7.5% acrylamide gels, then transferred to Whatman Protran nitrocellulose membranes. Ponceau-staining was performed to ensure protein transfer. The membrane was blocked in 5% skim milk in TBST for 1 hour at RT, followed by primary antibody incubation at 4 ºC in 2% skim milk solution in TBST overnight utilizing a polyclonal antibody raised against a section of the NLS region of NF B p50 (SC-114X, Santa Cruz). After washing the membrane, secondary incubation was performed with a goat anti-rabbit IgG-HRP antibody (Santa Cruz) in 2% skim milk solution in TBST for 1 hour at RT. The membrane was washed

again and incubated in SuperSignal West Pico Chemiluminescent Substrate (ThermoScientific) for 5 minutes before being exposed to x-ray film. K562 whole cell lysate (Santa Cruz) was used as a positive control for the NF B-western blot analysis. PARP (9542, Cell signaling) and β-tubulin (G098, abm) antibodies were used as loading controls for the nuclear and cytoplasmic fractions, respectively. For HTT protein analysis, primary skin fibroblasts were cultured as previously described [61] and cell pellets were lysed by resuspension in SDP buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1% Igepal, 40 mM B-glycerophosphate, 10 mM NaF, 1X Roche complete protease inhibitor, 1 mM sodium orthovanadate and 800 mM PMSF) containing 0.1% SDS. Protein samples were denatured in LDS sample buffer (Invitrogen) with 100mM DTT and incubated at 70°C for 10min. Samples were resolved on 10% low-BIS acrylamide gels (200:1 acrylamide:BIS) with tris-glycine running buffer (25mM Tris, 190mM glycine, 0.1% SDS) containing 10.7mM beta-mercaptoethanol added fresh using the BioRad Protean II xi Cell system with cooling unit. Gels were run at 160V for 60min through the stack, then 250V for 16h. Proteins were then transferred to nitrocellulose at 24V for 2h with NuPage transfer buffer (Invitrogen). Membranes were blocked with 5% milk in PBS, and then blotted for HTT with anti-HTT antibody (Millipore mAb2166) and anti-Myosin (Abcam ab24762), which was the loading control used to normalize the data. Proteins were detected with IR dye 800CW goat anti-mouse (Rockland 610-131-007) and AlexaFluor 680 goat anti-rabbit (Molecular Probes A21076) labeled secondary antibodies. The LiCor Odyssey Infrared Imaging system was used for signal detection.

siRNA transfection

HEK293 cells (Sigma 85120602) were seeded in 12-well plates at 2.5×10^5 cells/well. siRNA knock-down experiments were performed according to DharmaFECT 1 (#T-2001-02 GE Healthcare) Transfection Reagent manufacturer's protocols. Twenty-four hours after seeding, the cells were transiently transfected with control or NFKB-targeted siRNA for RELA (Smartpool 5970) and NFKB1 (Smartpool 4790) combined (final concentration of 50nM for each siRNA). Cells were collected for quantitative real-time PCR analysis (qPCR) after 72h treatment.

Quantitative real time PCR analysis

RNA was isolated from cells using the Qiagen RNeasy mini kit (Qiagen). 1000 ng of RNA was used for RT-PCR carried out with the Applied Biosystems High Capacity cDNA Reverse Transcription kit (Life Technologies). qPCR was performed with Invitrogen's Fast SYBR® Green Master Mix according to manual (Life Technologies) on the 7500 Fast Real-Time PCR System (Applied Biosystems). Absolute quantification was used and analysis was performed using the 7500 Software V2.0.1 (Applied Biosystems). A normalization factor calculated on ACTB and HPRT1 results was used to normalize the data. Primers used are described in Supplementary information.

HD patient materials

The HD patients analyzed in the Danish cohort originated from 36 Danish HD families, and have previously been described[36]. The B-haplotype families (n=8) in Denmark are likely to be connected via a common ancestor, an HD founder carrying the B-haplotype, as their chromosome 4p16.3 haplotypes are similar, and the families all originate from

the same geographical area. The HD patients genotyped in the UBC cohort were primarily Canadians of European origin obtained from the Hayden lab and the UBC HD BioBank (UBC CREB H06-70467 and H05-70532). To enable identification of disease-modifying SNP variants, the UBC cohort consisted of samples collected from a single member from each family with the most extreme AO phenotype if present as defined by the Langbehn *et al* formula.[8,37] In families with neither mean nor extreme outliers in terms of AO, a single randomly selected family member was used. The EHDN Registry cohort represented a random cohort with patient data obtained from sixteen different countries in Europe (with multiple sites within each country) with no specific selection strategies applied for inclusion. Ethical approval was obtained from the local ethics committee for each study site. HD subjects were categorized into percentiles based on their ratio AO (observed AO/expected AO); Mean (40-60th) percentile (ratio AO=0.93-1.05); early 15-40th percentile (ratio AO=0.93-0.83); extremely early <15th percentile (ratio AO<0.83); late 60-85th percentile (ratio AO=1.05-1.17); extremely late >85th percentile (ratio AO>1.17. AO reporting procedures differed between the cohorts. The EHDN Registry cohort involved historical self-reporting, whereas the same clinicians assessed the patients in the UBC and Danish cohorts longitudinally.

Exclusion of HD patients in cohort data

In the Danish cohort the total number of genotyped HD subjects were n=98: 36 (G/A) carriers with the (A) variant phased to the HD disease allele, three (G/A) carriers with the (A) variant phased to the wild-type allele, 53 (G/G) carriers and six (A/A) homozygotes. Due to the very low number the three (G/A) heterozygous patients in the Danish cohort with the (A) variant phased to the wild-type allele were excluded from analysis. Subjects

in the Danish cohort were excluded with CAG repeat lengths of 40 or less (n=6) due to clear evidence of biased observation towards those with early onset. Two subjects with CAG length 41 and genotype (G/G) were also excluded due to extremely early reported age of onset with 22 and 35 years prior to expected onset. This bias is consistent with the prior experience of Langbehn *et al* [8]. Note that exclusion of these subjects decreases the estimated protective effect of the (A) sequence variant when in phase with the HD disease allele. Five (G/G) homozygotes with CAG lengths greater than 55 were also excluded. They add no evidence to the assessment of the SNP effect, their CAG-expected age of onset is not well-established using the same analysis techniques used for the CAG range 41-55, and their early onset ages only allows them to serve as potentially influential but irrelevant outliers. In the UBC cohort the total number genotyped HD subjects were n=459: 26 (G/A) carriers with the (A) variant phased to the wild-type allele, 2 (G/A) carriers with the (A) variant phased to the HD disease allele and 431 (G/G) carriers. In the UBC cohort, four (G/A) heterozygotes and 40 (G/G) homozygotes with CAG lengths lesser than 41 or greater than 55 were excluded from primary analysis. Two (G/A) subjects with the (A) variant phased on the HD disease allele were also excluded from the UBC data analysis. In the EHDN Registry cohort, the total number of genotyped HD subjects including all CAG repeat lengths were (n=497): 32 (G/A) carriers with the (A) variant phased to the wild-type allele, 459 (G/G) carriers and 6 (G/A) carriers with the (A) variant phased to the HD disease allele. In the EHDN Registry cohort, six (G/A) heterozygotes with the (A) variant phased on the wild-type allele and 41 (G/G) homozygotes with CAG lengths lesser than 41 or greater than 55 were excluded from primary analysis as well as the six (G/A) heterozygotes with the (A) variant phased

on the HD disease allele. There was no evidence for biased ascertainment at CAG repeat length 40 in the UBC and the EHDN Registry cohorts. However, for comparability and uniformity of analysis, all three data sets *i.e* the UBC, EHDN Registry and the Danish cohorts, were analyzed using the same cutoff of CAG lengths 41-55. Analysis was also performed including HD subjects with CAG repeat length 40 in the UBC and the EHDN Registry data set with no substantive changes in the results (data not shown).

Statistical analysis of clinical data

For analysis of the clinical human data, we studied the relationship between HD age of onset and the presence and phase of the human SNP rs13102260. For the Danish data, we fit weighted least square models predicting observed ages of onset as a function of the rs13102260 (G/A) genotype, and expected age of onset based on CAG length, as estimated by Langbehn *et al* [8]. We controlled for a random effect of family pedigree. Because variance in age of onset tends to increase with higher expected age of onset (lower CAG length), we also included a random effect for expected age of onset, effectively turning this into a weighted least squares model based on CAG-predicted onset age. Analyses were performed using PROC Mixed of SAS 9.3, and the Kenward-Rogers procedure was used to estimate residual degrees of freedom. We used similar models for the UBC and Registry data, except that all subjects were from separate families in the UBC cohort and possible familial relationships were unknown in the EHDN Registry cohort. Thus, no random family effect could be used. Because the UBC data were collected on the basis of age of onset, with SNP genotype as the unknown variable being predicted, we also compared the linear model result to those from the corresponding logistic regression predicting SNP genotype as the outcome. The

difference between the observed and CAG-based expected onset age was the main predictor of allele frequency. For comparability to the mixed linear models, we also include expected age of onset as a covariate, though it lacked significance as a predictor of rs13102260 (G/A) genotype.

SNP genotyping

The Taqman® SNP Genotyping assay (C_31758132_10) (Applied Biosystems) was used for genotyping of the rs13102260 (G>A). TaqMan® Genotyping master mix (Applied Biosystems) was used and the reaction was set up according to manufacturer's recommendations. PCR conditions: 95°C for 10 min; 92°C for 15 sec and anneal/extend at 62°C for 1 min for 40 cycles using the 7500 Fast Real-Time PCR system (Applied Biosystems). Analysis was performed using the 7500 Software V2.0.1 (Applied Biosystems). In the Danish cohort the rs13102260 (G/A) genotype was assessed using Sanger sequencing using the following primer sequences: Forward 5'-GCCTCACCCCATTACAGTCT-3' and Reverse 5'-GGCAATGAATGGGGCTCT-3'

Phasing of (G/A) carriers

In the Danish cohort phasing of the rs13102260 (G/A) genotype was done in each HD family by segregation of affected and/or unaffected genotypes within the pedigree, similar to the procedure applied in the UBC cohort. In the UBC cohort, all HD subjects were first genotyped for rs13102260. We then genotyped relatives of the heterozygous (G/A) carriers to phase the (A) variant to the wild-type or the HD disease allele. Relatives included asymptomatic or unaffected offspring, parents or siblings, or affected distant relatives such as cousins, nieces and nephews The unaffected parent was almost

exclusively (G/A) and the affected parent was (G/G) (26 out of 28). The (A) variant thus segregated with the wild-type allele inherited from the unaffected parent for all the (G/A) carriers that were included in the association analysis. For (G/A) carriers in the EHDN Registry we used a PCR-based approach for phasing. Human HD patient DNA was PCR-amplified using the following primers: forward 5'-ATTACAGTCTCACCACGCCC-3'; reverse 5'-GACAAGGGAAGACCCAAGTG-3'. The wild-type and the CAG-expanded HD disease alleles were separated by running the PCR product on a 0.7% agarose gel. DNA was extracted using a gel extraction kit (Invitrogen) and DNA concentrations were determined using the Nanodrop spectrophotometer (ThermoScientific). Capillary sequencing (Applied Biosystems) was performed using the same primers as for the initial PCR amplification. The obtained sequences incorporated the rs13102260 (G/A) and the CAG tract. Sequence Scanner v1.0 (Applied Biosystems) and ApE Plasmid editor software were used for sequence analysis.

Statistical analysis of non-clinical data

Student's unpaired t-test was performed for comparison between two groups (GraphPad Prism5). One-way analysis of variance (ANOVA) with Bonferroni's or Dunnett's multiple comparison tests were used for data analysis of more than two groups (GraphPad Prism5). Significant differences were set at *P<0.05, **P<0.01, ***P<0.001.

ACKNOWLEDGMENTS

The authors wish to thank Michael Davey and Nicoletta Schintu for technical assistance, Rebecca Hunt Newbury for helpful discussions of the manuscript, Joakim Tedroff for critical assessment of clinical data, Robert A. Harris for critical reading of the manuscript and Lisbeth Vinding Olsen for excellent technical assistance.

AUTHOR CONTRIBUTIONS

K.B conceived, designed and performed the experiments, analyzed and interpreted the data, and wrote the manuscript. T.L., S.J.N., S.M., G.G., and J.B. performed experiments. A.N. performed SNP genotyping and haplotype identification of Danish cohort, contributed with fibroblast lines, and to writing the manuscript C.K., J.A.C., C.D., and S.C.W. conceived and conducted haplotype identification experiments of UBC cohort. D.A. and E.P.C. performed *in silico* analysis. N.L. and S.J.T., contributed with DNA and conducted SNP genotyping of EHDN Registry samples. C.C. and R.A.G.DS contributed with cell culture work. O.H contributed to ChIP and siRNA experiments and to writing the manuscript. M.R.H provided the UBC HD biobank samples. D.R.L. performed statistical analysis of cohort data. W.W.W conceived experiments. B.R.L conceived experiments and contributed to writing manuscript. This work was supported by the Swedish Research Council (K.B), Canadian Institutes of Health Research (B.R.L., M.R.H., W.W.W.), CHDI foundation (K.B., M.R.H., B.R.L). M.R.H. is a Canadian Research Chair in Molecular Medicine.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. Strong, T.V. *et al.* Widespread expression of the human and rat Huntington's disease gene in brain and nonneural tissues. *Nat.Genet.* **5**, 259-265 (1993).
2. Li, S.H. *et al.* Huntington's disease gene (IT15) is widely expressed in human and rat tissues. *Neuron* **11**, 985-993 (1993).
3. Sotrel, A. *et al.* Morphometric analysis of the prefrontal cortex in Huntington's disease. *Neurology* **41**, 1117-1123 (1991).
4. Leavitt, B.R. *et al.* Wild-type huntingtin protects neurons from excitotoxicity. *J Neurochem* **96**, 1121-9 (2006).
5. Cattaneo, E. *et al.* Loss of normal huntingtin function: new developments in Huntington's disease research. *Trends Neurosci* **24**, 182-8 (2001).
6. Gil, J.M. & Rego, A.C. Mechanisms of neurodegeneration in Huntington's disease. *Eur J Neurosci* **27**, 2803-20 (2008).
7. Andrew, S.E. *et al.* The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease [see comments]. *Nat.Genet.* **4**, 398-403 (1993).
8. Langbehn, D.R., Brinkman, R.R., Falush, D., Paulsen, J.S. & Hayden, M.R. A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. *Clin Genet* **65**, 267-77 (2004).
9. Brinkman, R.R., Mezei, M.M., Theilmann, J., Almqvist, E. & Hayden, M.R. The likelihood of being affected with Huntington disease by a particular age, for a specific CAG size. *Am.J.Hum.Genet.* **60**, 1202-1210 (1997).

10. Zabetian, C.P. *et al.* A quantitative-trait analysis of human plasma-dopamine beta-hydroxylase activity: evidence for a major functional polymorphism at the DBH locus. *Am J Hum Genet* **68**, 515-22 (2001).

11. Rigat, B. *et al.* An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. *J Clin Invest* **86**, 1343-6 (1990).

12. Wexler, N.S. *et al.* Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proc Natl Acad Sci U S A* **101**, 3498-503 (2004).

13. Van Raamsdonk, J.M. *et al.* Body weight is modulated by levels of full-length huntingtin. *Hum Mol Genet* **15**, 1513-23 (2006).

14. Graham, R.K. *et al.* Levels of mutant huntingtin influence the phenotypic severity of Huntington disease in YAC128 mouse models. *Neurobiol Dis* **21**, 444-55 (2006).

15. Leavitt, B.R. *et al.* Wild-type huntingtin reduces the cellular toxicity of mutant huntingtin in vivo. *Am J Hum Genet* **68**, 313-324 (2001).

16. Van Raamsdonk, J.M. *et al.* Loss of wild-type huntingtin influences motor dysfunction and survival in the YAC128 mouse model of Huntington disease. *Hum Mol Genet* **14**, 1379-92 (2005).

17. Coles, R., Caswell, R. & Rubinsztein, D.C. Functional analysis of the Huntington's disease (HD) gene promoter. *Hum.Mol.Genet.* **7**, 791-800 (1998).

18. Holzmann, C. *et al.* Isolation and characterization of the rat huntingtin promoter. *Biochem.J.* **336 (Pt 1)**, 227-234 (1998).

19. Feng, Z. *et al.* p53 tumor suppressor protein regulates the levels of huntingtin gene expression. *Oncogene* **25**, 1-7 (2006).

20. Wang, R. *et al.* Sp1 regulates human huntingtin gene expression. *J Mol Neurosci* **47**, 311-21 (2012).

21. Cohen, J.C. *et al.* Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**, 869-72 (2004).

22. Cohen, J. *et al.* Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet* **37**, 161-5 (2005).

23. Johansen, C.T. *et al.* Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* **42**, 684-7 (2010).

24. Emond, M.J. *et al.* Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. *Nat Genet* **44**, 886-9 (2012).

25. Ott, J., Kamatani, Y. & Lathrop, M. Family-based designs for genome-wide association studies. *Nat Rev Genet* **12**, 465-74 (2011).

26. Barnett, I.J., Lee, S. & Lin, X. Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. *Genet Epidemiol* **37**, 142-51 (2013).

27. Hayles, B., Yellaboina, S. & Wang, D. Comparing transcription rate and mRNA abundance as parameters for biochemical pathway and network analysis. *PLoS One* **5**, e9908 (2010).

28. Hodgson, J.G. *et al.* Human huntingtin derived from YAC transgenes compensates for loss of murine huntingtin by rescue of the embryonic lethal phenotype. *Hum.Mol.Genet.* **5**, 1875-1885 (1996).

29. Cattaneo, E. & Conti, L. Generation and characterization of embryonic striatal conditionally immortalized ST14A cells. *J Neurosci Res* **53**, 223-34 (1998).

30. Ehrlich, M.E. *et al.* ST14A cells have properties of a medium-size spiny neuron. *Exp Neurol* **167**, 215-26 (2001).

31. Andersen, M.C. *et al.* In silico detection of sequence variations modifying transcriptional regulation. *PLoS Comput Biol* **4**, e5 (2008).

32. Chen, F.E., Huang, D.B., Chen, Y.Q. & Ghosh, G. Crystal structure of p50/p65 heterodimer of transcription factor NF-kappaB bound to DNA. *Nature* **391**, 410-3 (1998).

33. Escalante, C.R., Shen, L., Thanos, D. & Aggarwal, A.K. Structure of NF-kappaB p50/p65 heterodimer bound to the PRDII DNA element from the interferon-beta promoter. *Structure* **10**, 383-91 (2002).

34. Grunberger, D. *et al.* Preferential cytotoxicity on tumor cells by caffeic acid phenethyl ester isolated from propolis. *Experientia* **44**, 230-2 (1988).

35. Natarajan, K., Singh, S., Burke, T.R., Jr., Grunberger, D. & Aggarwal, B.B. Caffeic acid phenethyl ester is a potent and specific inhibitor of activation of nuclear transcription factor NF-kappa B. *Proc Natl Acad Sci U S A* **93**, 9090-5 (1996).

36. Norremolle, A. *et al.* 4p16.3 haplotype modifying age at onset of Huntington disease. *Clin Genet* **75**, 244-50 (2009).

37. Langbehn, D.R., Hayden, M.R. & Paulsen, J.S. CAG-repeat length and the age of onset in Huntington disease (HD): a review and validation study of statistical approaches. *Am J Med Genet B Neuropsychiatr Genet* **153B**, 397-408 (2009).

38. Andrew, S.E., Goldberg, Y.P., Theilmann, J., Zeisler, J. & Hayden, M.R. A CCG repeat polymorphism adjacent to the CAG repeat in the Huntington disease gene: implications for diagnostic accuracy and predictive testing. *Hum Mol Genet* **3**, 65-7 (1994).

39. Pecheux, C. *et al.* Sequence analysis of the CCG polymorphic region adjacent to the CAG triplet repeat of the HD gene in normal and HD chromosomes. *J Med Genet* **32**, 399-400 (1995).

40. Coles, R., Leggo, J. & Rubinsztein, D.C. Analysis of the 5' upstream sequence of the Huntington's disease (HD) gene shows six new rare alleles which are unrelated to the age at onset of HD. *J.Med.Genet.* **34**, 371-374 (1997).

41. Nadeau, J.H. Modifier genes in mice and humans. *Nat Rev Genet* **2**, 165-74 (2001).

42. Genin, E., Feingold, J. & Clerget-Darpoux, F. Identifying modifier genes of monogenic disease: strategies and difficulties. *Hum Genet* **124**, 357-68 (2008).

43. Gusella, J.F. & MacDonald, M.E. Huntington's disease: the case for genetic modifiers. *Genome Med* **1**, 80 (2009).

44. Metzger, S. *et al.* Huntingtin-associated protein-1 is a modifier of the age-at-onset of Huntington's disease. *Hum Mol Genet* **17**, 1137-46 (2008).

45. Naze, P., Vuillaume, I., Destee, A., Pasquier, F. & Sablonniere, B. Mutation analysis and association studies of the ubiquitin carboxy-terminal hydrolase L1 gene in Huntington's disease. *Neurosci Lett* **328**, 1-4 (2002).

46. Weydt, P. *et al.* The gene coding for PGC-1alpha modifies age at onset in Huntington's Disease. *Mol Neurodegener* **4**, 3 (2009).

47. Metzger, S. *et al.* Age at onset in Huntington's disease is modified by the autophagy pathway: implication of the V471A polymorphism in Atg7. *Hum Genet* (2010).

48. Ge, B. *et al.* Global patterns of cis variation in human cells revealed by high-density allelic expression analysis. *Nat Genet* **41**, 1216-22 (2009).

49. Kasowski, M. *et al.* Variation in transcription factor binding among humans. *Science* **328**, 232-5 (2010).

50. Meffert, M.K. & Baltimore, D. Physiological functions for brain NF-kappaB. *Trends Neurosci* **28**, 37-43 (2005).

51. Van Raamsdonk, J.M., Pearson, J., Murphy, Z., Hayden, M.R. & Leavitt, B.R. Wild-type huntingtin ameliorates striatal neuronal atrophy but does not prevent other abnormalities in the YAC128 mouse model of Huntington disease. *BMC Neurosci* **7**, 80 (2006).

52. Liu, W. *et al.* Increased Steady-State Mutant Huntingtin mRNA in Huntington's Disease Brain. *J Huntingtons Dis* **2**, 491-500 (2013).

53. Zhang, Y., Engelman, J. & Friedlander, R.M. Allele-specific silencing of mutant Huntington's disease gene. *J Neurochem* **108**, 82-90 (2009).

54. Boudreau, R.L. *et al.* Nonallele-specific silencing of mutant and wild-type huntingtin demonstrates therapeutic efficacy in Huntington's disease mice. *Mol Ther* **17**, 1053-63 (2009).

55. Drouet, V. *et al.* Sustained effects of nonallele-specific Huntingtin silencing. *Ann Neurol* **65**, 276-85 (2009).

56. DiFiglia, M. *et al.* Therapeutic silencing of mutant huntingtin with siRNA attenuates striatal and cortical neuropathology and behavioral deficits. *Proc Natl Acad Sci U S A* **104**, 17204-9 (2007).

57. Harper, S.Q. *et al.* RNA interference improves motor and neuropathological abnormalities in a Huntington's disease mouse model. *Proc Natl Acad Sci U S A* **102**, 5820-5 (2005).

58. Carroll, J.B. *et al.* Potent and selective antisense oligonucleotides targeting single-nucleotide polymorphisms in the Huntington disease gene / allele-specific silencing of mutant huntingtin. *Mol Ther* **19**, 2178-85 (2011).

59. Yu, D. *et al.* Single-stranded RNAs use RNAi to potently and allele-selectively inhibit mutant huntingtin expression. *Cell* **150**, 895-908 (2012).

60. Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* **32**, D91-4 (2004).

61. Norremolle, A. *et al.* Mosaicism of the CAG repeat sequence in the Huntington disease gene in a pair of monozygotic twins. *Am J Med Genet A* **130a**, 154-9 (2004).

FIGURE LEGENDS

Figure 1 *Cis*-regulatory SNP in the NFκB binding site alters transcriptional activity of the *HTT* promoter.

Reporter assays measuring basal transcriptional activity of twelve *HTT* promoter constructs. The transcriptional activity of construct 5 was reduced by (a) 57% (P<0.01, two-tailed unpaired t-test) in HEK293A cells and by (b) 51% (P<0.001) in ST14A cells compared to its parent construct 4. (c) Cartoon of the *HTT* promoter regions (A, B, C) that were investigated for regulation of expression activity. Construct 4 differed from construct 5 at ten sequence positions as indicated (relative to the translation start site). *In silico* analysis identified putative transcriptional factor binding sites (TFBS) potentially affected by the sequence variant differences. Genetic variants that overlap with putative TFBS were identified. Validated SNPs are indicated with blue triangle. Variants without frequency submission are indicated in gray. (d) The proximal promoter fragment C from construct 5 reduced transcriptional activity of the construct 4 promoter background, while the construct 4 fragment C rescued the transcriptional activity of construct 5 (P<0.001). (e) The 5 C fragment alone was sufficient for the reduced transcriptional activity (P<0.001). (f) Sequence variation in the NF-κB TFBS caused the reduced transcriptional activity observed in the construct 5 full-length and fragment C constructs (P<0.001). (g) TFBS motifs and respective construct 4 and 5 sequence variants. The identified NFκB TFBS in the *HTT* promoter comprises the rs13102260 (G>A). (h) Site-directed mutagenesis changing G□A in the first position of the NF-κB TFBS in construct 4 C reduced transcriptional activity to similar levels as observed for the construct 5 full-length and C-fragment (P<0.001). Each construct was assayed in triplicate. The data is

pooled from two independent experiments (mean ± SEM, n=6) (One-way ANOVA with with Bonferroni's multiple comparison test was used in (d) and Dunnett's multiple comparison test was used in (e-h).

Figure 2 NF-┤B binds to the huntingtin promoter *in vitro* and *in vivo*

(a-c) ChIP on different brain regions of naïve mice was used to investigate whether NF-┤B bound the site *in vivo.* The mouse *Htt* promoter constitutes three putative NF-┤B TFBS within 1000 bp upstream of the translation initiation codon. There was a NF-┤B enrichment at all analyzed TFBS regions comparable to the positive control region for IL6 as assessed by qPCR (Striatum $F(7,32)=1.75$, $P=0.133$; Prefrontal cortex $F(7,32)=2.44$, $P<0.05$, Cerebellum $F(7,32)=7.22$, $P<0.001$, one-way ANOVA) (mean ± SEM, n=5). (d) There were significant increases in NF-┤B recruitment in striatum compared to pre-frontal cortex and cerebellum at TFBS2 (Tissue $F(2,96)=5.53$, $P<0.01$, TFBS $F(7,96)=3.15$, $P<0.01$, interaction ($F(14,96)=4.28$, $P=0.311$, two-way ANOVA). (e) To study the NF-┤B occupancy at the *HTT* promoter in a human context, ChIP analysis was performed in lymphoblastoid cell lines (LCL) derived from HD patients. There was an increased enrichment of NF-┤B at the predicted NF-┤B TFBS containing the rs13102260 in human LCLs ($F(3,32)=4,28$, $P<0.05$, one-way ANOVA with Dunnett's test) compared to a control region upstream in the *HTT* promoter region containing two putative NF-┤B TFBS (-2065; -2011bp upstream from translation start site) (mean ± SEM, n=9). (f) Rat striatal ST14A cells stimulated with TNFα were assessed. Four putative NF-┤B TFBS were identified in the rat *Htt* promoter within 1000 bp upstream of the translation initiation codon (TFBS1 containing two predicted binding sites). We

observed an enrichment of NF-┤ B for all three regions comparable to what was observed for the IL6 promoter ($F(7,16)=5.09$, $P<0.01$, one-way ANOVA with Bonferroni's test) (mean ± SEM, n=3). Immunoprecipitation with a normal mouse IgG antibody was used as a negative control.

Figure 3 *Cis*-regulatory variants in the *HTT* promoter alter NF-κB binding

(a,b) Electromobility shift asays (EMSA) were performed to assess the binding of NF-κB to the putative NF-κB TFBS identified in constructs 4 and 5. Binding of nuclear extracts from TNFα-stimulated ST14A cells and recombinant NF-κB p50 protein were tested. NF-κB bound stronger to the allele 4 compared to the allele 5 oligonucleotide in a concentration-dependent manner. EMSA showed that NF-┤ B binding to the oligonucleotide containing the rs13102260 (A) variant was fully abolished, as assessed with both (c) ST14A nuclear extract and (d) recombinant NF-┤ B p50 protein. Lanes 1-2, no nuclear extract; lanes 3, 5, 7, allele 4 oligonucleotide and lanes 4, 6, 8, allele 5 oligonucleotide incubated with ST14A nuclear extract or recombinant NF-κB p50, respectively; lanes 9-10, unlabeled competitor oligonucleotide added to labeled oligonucleotide + protein extract. (gsu=gel shift units).

Figure 4 Targeting of NF-κB modulates *HTT* expression

(a) siRNA knock-down of NF-┤ B in HEK293 cells. Down-regulation of NFKB1 (specific for p105, precursor to the p50 subunit) and p65 mRNA levels, decreased the levels of *HTT* mRNA expression as measured by qPCR ($t(6)=3.53$, $P<0.01$) (mean ± SEM, n=7). siRNA knockdown efficiency was on average 83% for NFKB1 ($t(16.8)$, $p<0.0001$) and 70% for p65 ($t(6)=7.95$, $P<0.0001$). (b) ST14A cells expressing constructs 4 or 5 were

stimulated with TNFα. The TNFα-stimulation resulted in increased transcriptional activity of construct 4 compared to the untreated control ($F(3,8)=29.7$, $P<0.001$, one-way ANOVA with Dunnett's test). No increase in transcriptional activity was detected for construct 5 upon TNFα-stimulation. (c) CAPE treatment of ST14A cells expressing construct 4 reduced the TNFα-induced increase in transcriptional activity in a dose-dependent manner by inhibition of NFκB activity ($F(4,7)=6.24$, $P<0.05$, one-way ANOVA with Dunnett's test). One representative data set out of three independent reporter assay experiments is shown for each panel (mean ± SEM). (d) TNFα-stimulation of the ST14A cells resulted in increased protein levels of NFκB subunits as measured by western blot analysis. The p50 subunit increased the most in the nuclear fraction following TNFα-stimulation of ST14A cells. The p65 (65 kDa), RelB (68 kDa) and c-Rel (69 kDa) were also increased, but not to the same extent as p50. The cytoplasmic fraction showed strong bands for the precursor subunits p100/p105, in addition to bands for the p50 subunit, which were expressed at similar levels across the different treatments. CAPE treatment had no effect on expression of the NF-κB protein. PARP and β-tubulin show equal loading of protein across lanes for the nuclear and cytoplasmic fractions, respectively**.**

Figure 5 rs13102260 (A) variant alters NF-κB binding and modulates HD age of onset. (a) Familial HD cases from Denmark were genotyped for the rs13102260. Presence of the rs13102260 (A) variant on the HD disease allele had a clear association with later age of HD onset (35 of 82 HD subjects were (G/A) heterozygotes) ($t=5.08$, 27.1 df, $P<0.0001$). (b) HD patients from the UBC BioBank cohort were genotyped for the rs13102260 (G/A). Presence of the rs13102260 (A) variant on the wild-type allele

was associated with an earlier age of HD onset (22 of 413 HD subjects were (G/A) heterozygotes) (P= t=-2.57, df=367, P=0.010). (c) HD cases (Danish and UBC cohorts) were categorized according to the AO percentile, given their expected AO and the observed AO. A higher proportion of early AO cases was observed in (G/A) carriers with the (A) variant phased to wild-type allele (64%, of which 41% displayed extremely early AO below $<15^{th}$ percentile) compared to the (G/G) (41%, of which 21% displayed extremely early AO below $<15^{th}$ percentile), while a higher proportion of late AO cases was observed in (G/A) carriers (63%, of which 26% displayed extremely late AO above $>85^{th}$ percentile) with the (A) variant phased to the HD disease allele compared to (G/G) carriers (24%, of which 9% displayed extremely late AO above $>85^{th}$ percentile). (d) ratio AO (=observed AO/expected AO) showed the effect of rs13102260 on age of onset. HD subjects with the rs13102260 (A) variant phased to the wild-type allele showed lower ratio AO values, while HD subjects with the (A) variant phased to the HD disease allele showed higher ratio AO values, compared to the (G/G) carriers ((A) variant phased to wild-type allele n=22; (A) variant phased to the HD disease allele n=35; (G/G) carriers n=432).

Figure 6 Genotype at rs13102260 drives allele-specific expression of HTT and modulates HD age of onset

(a) The effect of the rs13102260 (A) variant on wild-type and mutant HTT (mHTT) protein expression levels was measured with western blot analysis in HD patient fibroblast lines. Wild-type HTT protein was reduced in the samples with the (A) variant phased to the wild-type allele compared to the (G) variant (G[WT] *vs* A[WT]) ($t(3)=3.27$, $P<0.05$, two-tailed unpaired t-test). Mutant HTT protein was reduced in the samples with

(A) phased on the HD disease allele compared to the (G) variant (G[HD] *vs* A[HD]) ($t(3)=2.46$, $P<0.05$). The data represents two different fibroblast lines per genotype group including two pellets per cell line (mean ± SEM, n=4). Alleles carrying the rs13102260 (A) variant are labelled in red; (G) variant carriers are indicated in white. For each genotype group, the left bar indicates wild-type HTT levels, and the right bar indicates mHTT levels. (b) Model for rSNP rs13102260 (G/A) as a bidirectional genetic modifier of HD age of onset. rSNP rs13102260 (G/A) is located in the identified NF-┤ B TFBS in the *HTT* promoter immediately proximal to the 5'UTR of the *HTT* gene. The rs13102260 (A) variant impairs NF-┤ B binding, reduces transcriptional activity of the *HTT* gene and HTT protein levels. Presence of the rs13102260 (A) variant on the wild-type allele was associated with reduced wild-type HTT protein levels and earlier AO, while the (A) variant on the HD disease allele was associated with lower mutant HTT protein levels and delayed AO in HD patients.

Figure 1

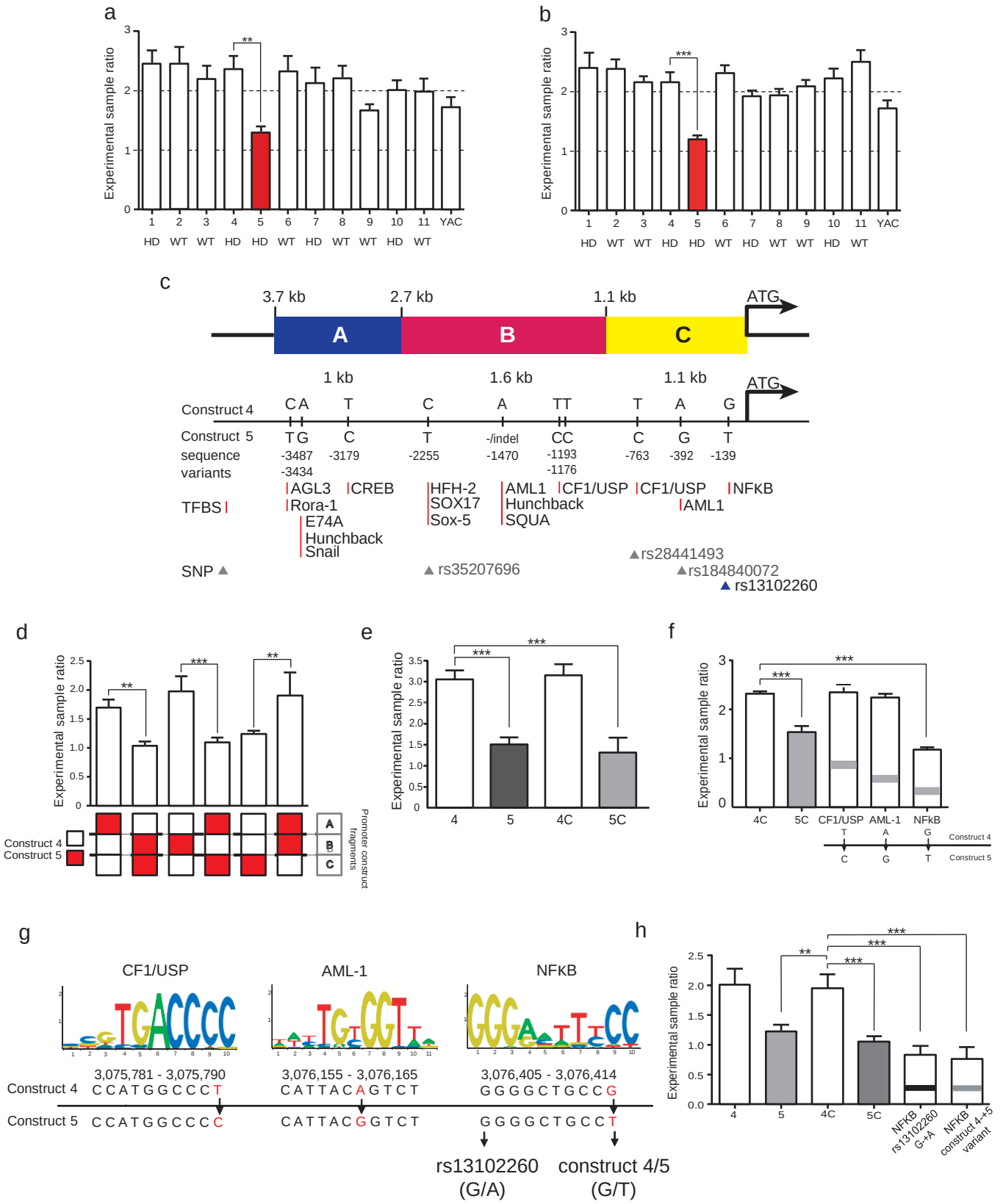

a
Experimental sample ratio
1 2 3 4 5 6 7 8 9 10 11 YAC
HD WT WT HD HD WT HD WT WT HD WT
b
Experimental sample ratio
1 2 3 4 5 6 7 8 9 10 11 YAC
HD WT WT HD HD WT HD WT WT HD WT
c
3.7 kb
2.7 kb
1.1 kb
ATG
A
B
C
1 kb
1.6 kb
1.1 kb
ATG
Construct 4
Construct 5
sequence variants
-3487
-3434
-3179
-2255
-1470
-1193
-1176
-763
-392
-139
TFBS
AGL3
CREB
Rora-1
E74A
Hunchback
Snail
HFH-2
SOX17
Sox-5
AML1
Hunchback
SQUA
CF1/USP
CF1/USP
AML1
NFκB
SNP
rs35207696
rs28441493
rs184840072
rs13102260
d
Experimental sample ratio
Construct 4
Construct 5
A
B
C
Promoter construct fragments
e
Experimental sample ratio
4
5
4C
5C
f
Experimental sample ratio
4C
5C
CF1/USP
AML-1
NFkB
Construct 4
Construct 5
g
CF1/USP
AML-1
NFκB
3,075,781 - 3,075,790
C C A T G G C C C T
3,076,155 - 3,076,165
C A T T A C A G T C T
3,076,405 - 3,076,414
G G G G C T G C C G
Construct 4
Construct 5
C C A T G G C C C C
C A T T A C G G T C T
G G G G C T G C C T
rs13102260 (G/A)
construct 4/5 (G/T)
h
Experimental sample ratio
4
5
4C
5C
NFKB rs13102260 G-A
NFKB construct 4-5 variant

Figure 2

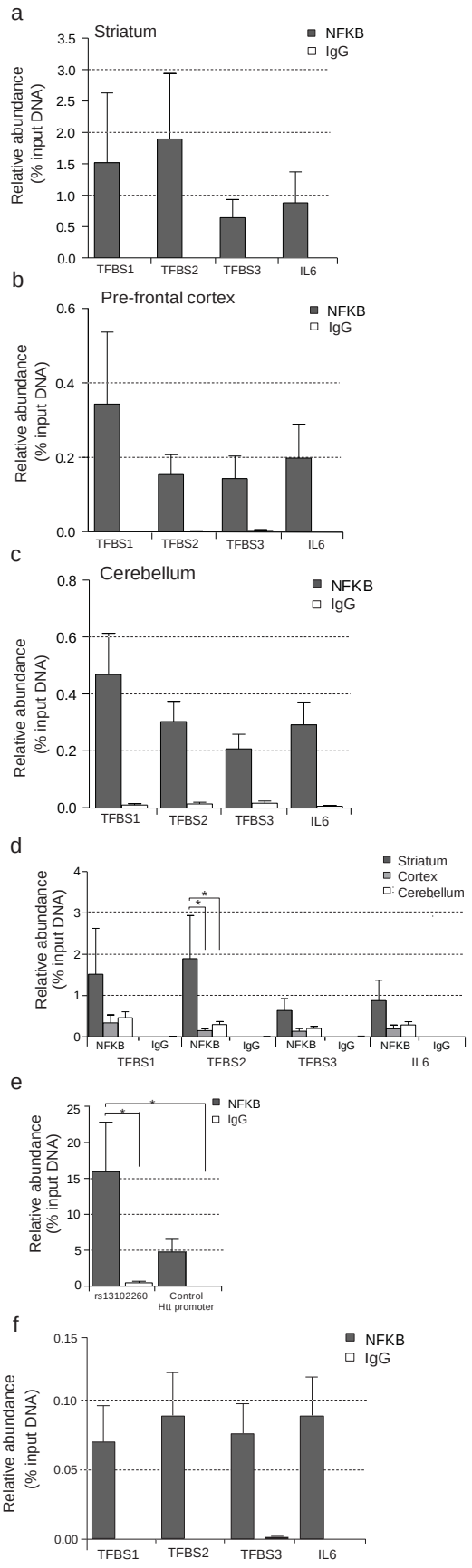

a
Striatum
NFKB
IgG
Relative abundance (% input DNA)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
TFBS1
TFBS2
TFBS3
IL6
b
Pre-frontal cortex
NFKB
IgG
Relative abundance (% input DNA)
0.6
0.4
0.2
0.0
TFBS1
TFBS2
TFBS3
IL6
c
Cerebellum
NFKB
IgG
Relative abundance (% input DNA)
0.8
0.6
0.4
0.2
0.0
TFBS1
TFBS2
TFBS3
IL6
d
Striatum
Cortex
Cerebellum
Relative abundance (% input DNA)
4
3
2
1
0
*
*
NFKB
IgG
NFKB
IgG
NFKB
IgG
NFKB
IgG
TFBS1
TFBS2
TFBS3
IL6
e
NFKB
IgG
Relative abundance (% input DNA)
25
20
15
10
5
0
*
*
rs13102260
Control Htt promoter
f
NFKB
IgG
Relative abundance (% input DNA)
0.15
0.10
0.05
0.00
TFBS1
TFBS2
TFBS3
IL6

Figure 3

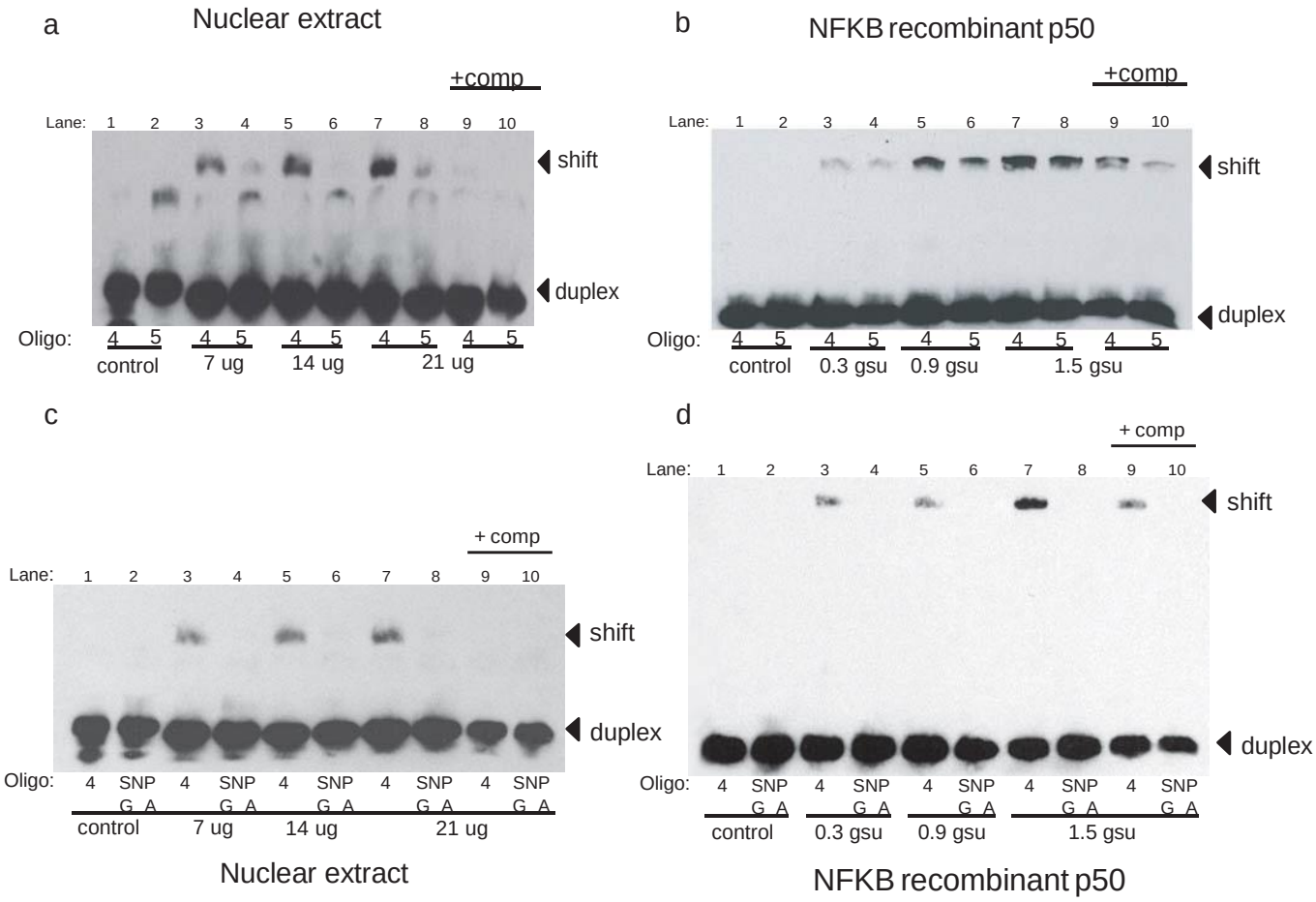

a
Nuclear extract
+comp
Lane: 1 2 3 4 5 6 7 8 9 10
shift
duplex
Oligo: 4 5 4 5 4 5 4 5 4 5
control
7 ug
14 ug
21 ug
b
NFKB recombinant p50
+comp
Lane: 1 2 3 4 5 6 7 8 9 10
shift
duplex
Oligo: 4 5 4 5 4 5 4 5 4 5
control
0.3 gsu
0.9 gsu
1.5 gsu
c
+ comp
Lane: 1 2 3 4 5 6 7 8 9 10
shift
duplex
Oligo: 4 SNP G A 4 SNP G A 4 SNP G A 4 SNP G A 4 SNP G A
control
7 ug
14 ug
21 ug
Nuclear extract
d
+ comp
Lane: 1 2 3 4 5 6 7 8 9 10
shift
duplex
Oligo: 4 SNP G A 4 SNP G A 4 SNP G A 4 SNP G A 4 SNP G A
control
0.3 gsu
0.9 gsu
1.5 gsu
NFKB recombinant p50

Figure 4

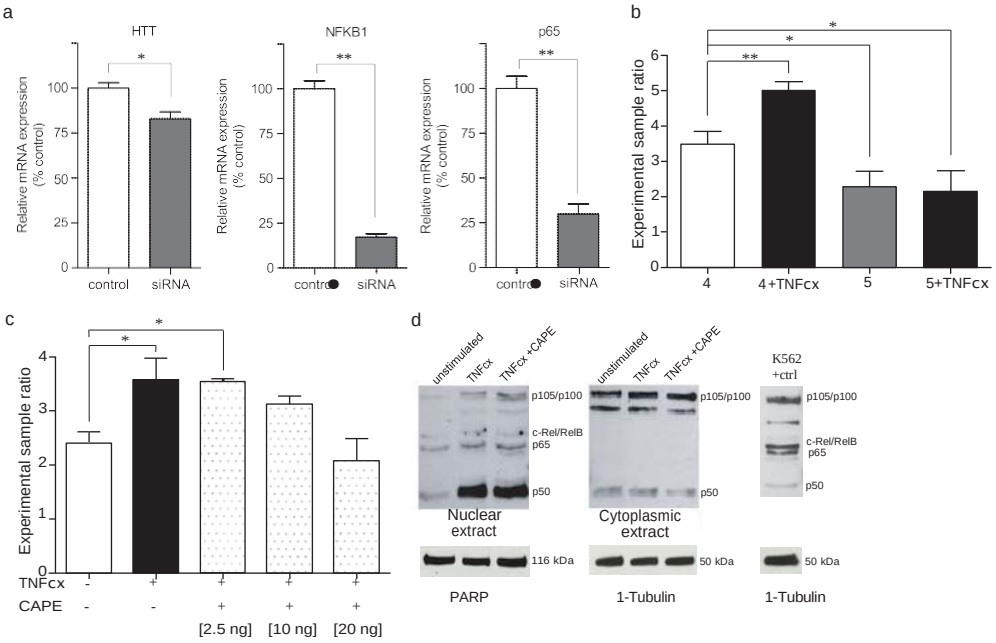
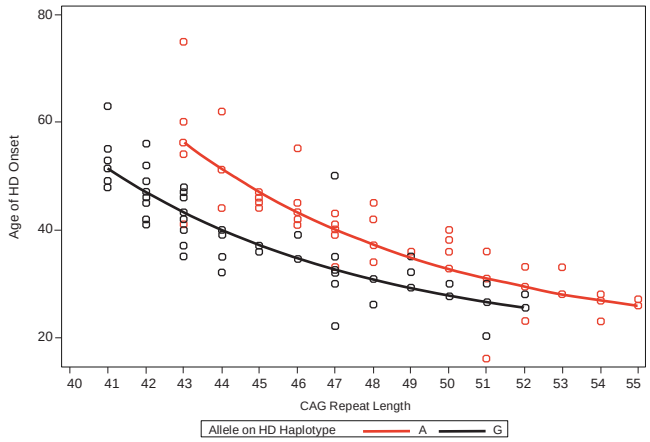
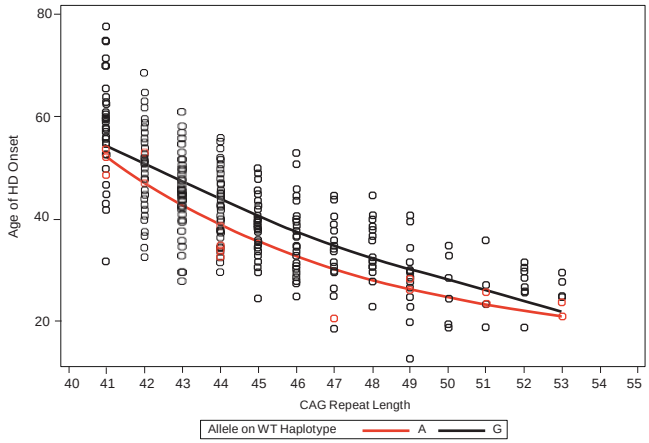

a
HTT
NFKB1
p65
Relative mRNA expression (% control)
control
siRNA
b
Experimental sample ratio
4
4+TNFcx
5
5+TNFcx
c
Experimental sample ratio
TNFcx
CAPE
[2.5 ng]
[10 ng]
[20 ng]
d
unstimulated
TNFcx
TNFcx +CAPE
p105/p100
c-Rel/RelB
p65
p50
Nuclear extract
Cytoplasmic extract
K562 +ctrl
116 kDa
50 kDa
PARP
1-Tubulin

Figure 5

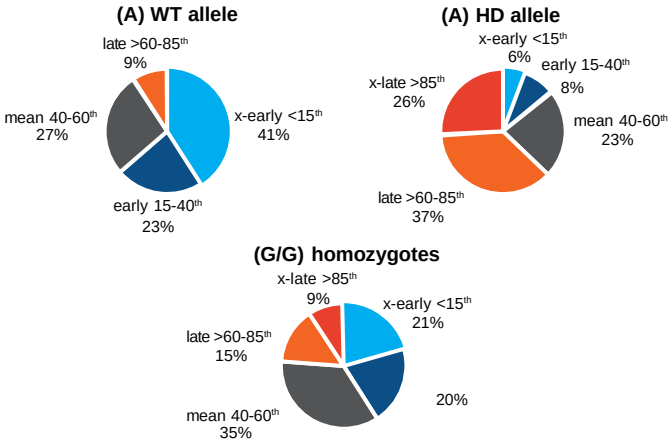
a



b



c



d

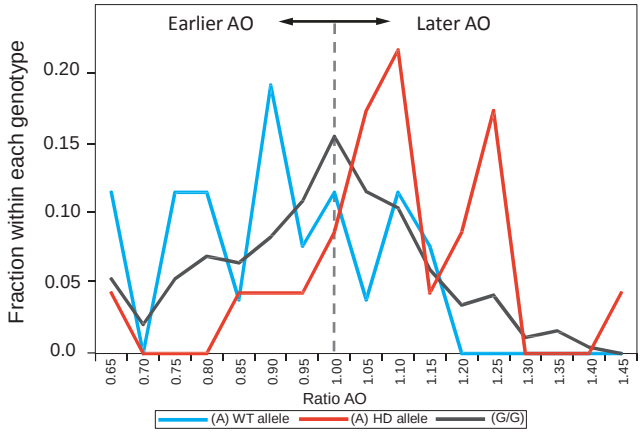
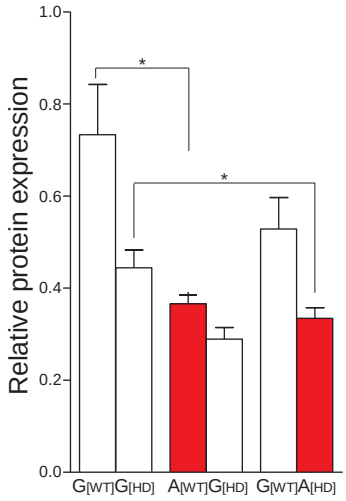


Figure 6

a


Relative protein expression
1.0
0.8
0.6
0.4
0.2
0.0
*
*
G[WT]G[HD]
A[WT]G[HD]
G[WT]A[HD]

b

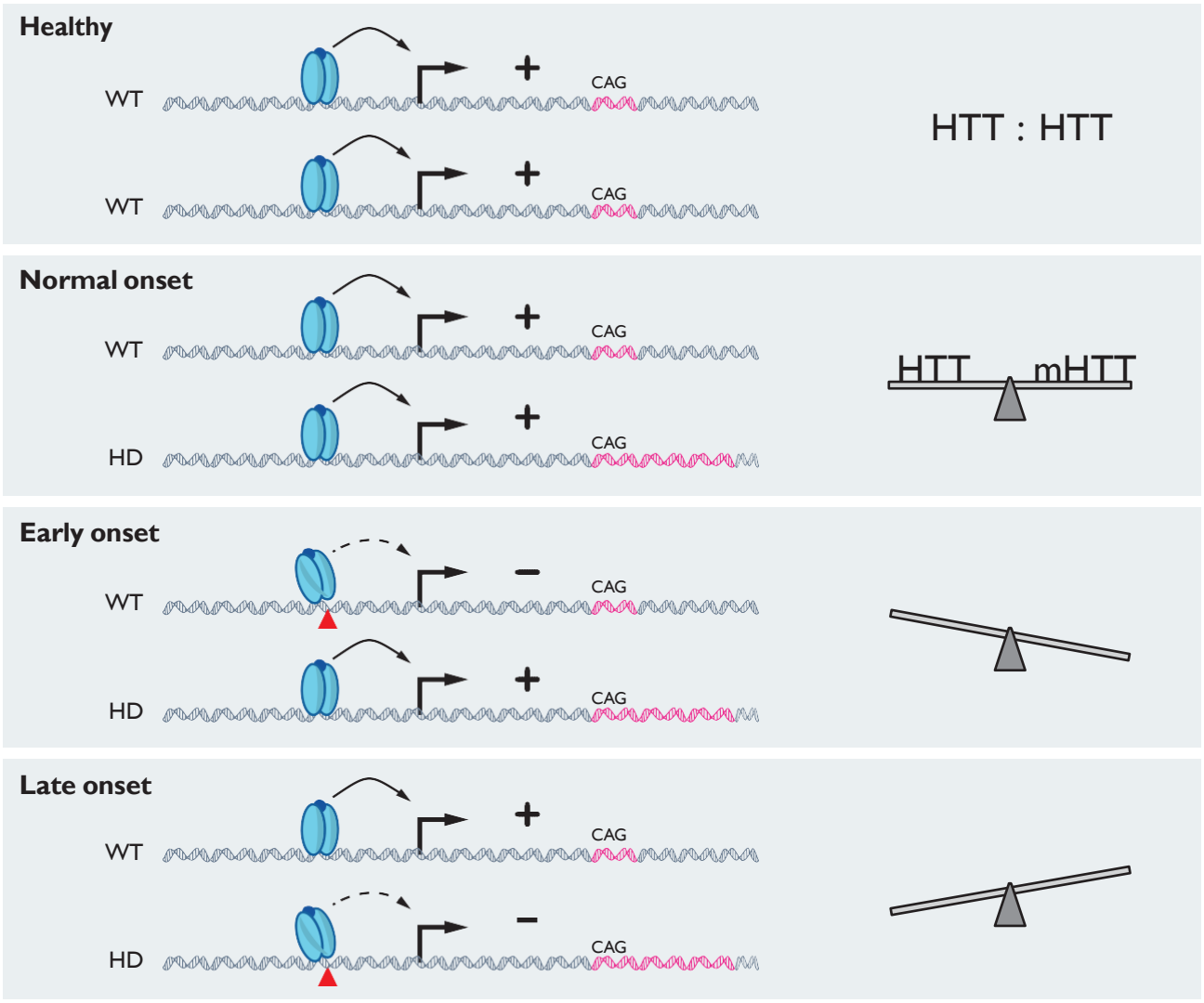

Healthy
WT
CAG
WT
CAG
HTT : HTT
Normal onset
WT
CAG
HD
CAG
HTT
mHTT
Early onset
WT
CAG
HD
CAG
Late onset
WT
CAG
HD
CAG

Table 1
Estimation of rs13102260 genotype effects on AO in HD subjects with (A) variant phased to HD disease allele

CAG length HD allele	Exp AO[1]	Model prediction (A)-HD allele-ABSENT[2]	Model prediction (A)-HD allele-PRESENT[3]	SD of ExpAO[1] (A)-HD allele[4]	SNP effect in years (A)-HD allele[5]
41	57.0	51.4	68.6	1.7	17.3
42	52.2	47.0	62.0	1.6	15.0
43	48.1	43.2	56.3	1.5	13.1
44	44.5	39.9	51.3	1.4	11.3
45	41.3	37.1	46.9	1.3	9.9
46	38.6	34.6	43.1	1.2	8.6
47	36.3	32.5	40.0	1.1	7.5
48	34.3	30.7	37.2	0.99	6.5
49	32.6	29.1	34.8	0.89	5.7
50	31.1	27.7	32.7	0.79	5.0
51	29.8	26.5	30.9	0.71	4.4
52	28.7	25.5	29.4	0.63	3.9
53	27.7	24.6	28.0	0.56	3.4
54	26.9	23.9	26.9	0.49	3.0
55	26.1	23.2	25.9	0.44	2.7

[1]Expected age of onset (Exp AO) calculated as defined by Langbehn *et al.* 2004
[2]Estimated AO based on model for rs13102260 (G) variant phased to HD disease allele.
[3]Estimated AO based on model for rs13102260 (A) variant phased to HD disease allele.
[4]Increase in standard deviation (SD) when rs13102260 (A) variant phased to HD disease allele.
[5]Estimated number of years delayed onset with rs13102260 (A) variant phased to HD disease allele.