A greatly extended **PPARGC1A** genomic locus encodes several new brain-specific isoforms and influences Huntington disease age of onset

Selma M. Soyal¹, Thomas K. Felder¹, Simon Auer¹, Penelope Hahne¹, Hannes Oberkofler¹, Anke Witting³, Markus Paulmichl², G. Bernhard Landwehrmeyer³, Patrick Weydt³ and Wolfgang Patsch¹,², *, For the European Huntington Disease Network ‡

¹Department of Laboratory Medicine, ²Department of Pharmacology and Toxicology, Paracelsus Medical University, Strubergasse 21, 5020 Salzburg, Austria and ³Department of Neurology, University of Ulm, Ulm, Germany

Received March 14, 2012; Revised and Accepted May 4, 2012

PGC-1α has been implicated in the pathogenesis of neurodegenerative disorders. Several single-nucleotide polymorphisms (SNPs) located in two separate haplotype blocks of **PPARGC1A** have shown associations with Huntington’s disease (HD) and Parkinson’s disease, but causative SNPs have not been identified. One SNP (rs7665116) was located in a highly conserved 233 bp region of intron 2. To determine whether rs7665116 is located in an alternative exon, we performed 5′-RLM-RACE from exon 3 and discovered multiple new transcripts that initiated from a common novel promoter located 587 kb upstream of exon 2, but did not contain the conserved region harboring rs7665116. Using real-time polymerase chain reaction, RNase protection assays and northern blotting, we show that the majority of these transcripts are brain specific and are at least equally or perhaps more abundant than the reference sequence **PPARGC1A** transcripts in whole brain. Two main transcripts containing independent methionine start codons encode full-length brain-specific PGC-1α proteins that differ only at their N-termini (NTs) from PGC-1α, encoded by the reference sequence. Additional truncated isoforms containing these NTs that are similar to NT-PGC-1α exist. Other transcripts may encode potential dominant negative forms, as they are predicted to lack the second LXXLL motif that serves as an interaction site for several nuclear receptors. Furthermore, we show that the new promoter is active in neuronal cell lines and describe haplotypes encompassing this region that are associated with HD age of onset. The discovery of such a large **PPARGC1A** genomic locus and multiple isoforms in brain warrants further functional studies and may provide new tissue-specific targets for treating neurodegenerative diseases.

**INTRODUCTION**

**PPARGC1A**, the gene encoding PGC-1α, was first cloned in mice (1), and homologues have since been cloned in species extending from *Drosophila* (2) to humans (3,4). The functions of this versatile transcriptional coregulator are extremely complex and range from mitochondrial biogenesis to glucose and lipid homeostasis in metabolically dynamic tissues such as liver, heart, adipose, brain and kidney (5–7).

An increasing body of experimental evidence implicates PGC-1α in neurodegenerative disorders (8,9). Independent groups presented evidence that PGC-1α plays a role in Huntington’s disease (HD) (10–12), amyotrophic lateral sclerosis (ALS) (13,14), Alzheimer’s disease (AD) (15) and Parkinson’s disease (PD) (16). In addition to enhancing mitochondrial biogenesis, PGC-1α also protects against an increased reactive oxygen species burden by inducing detoxifying enzymes (17,18). Thus, defective PGC-1α expression and/or function...
may contribute to the main pathogenetic features of neurodegenerative disorders such as transcriptional dysregulation, mitochondrial impairment and increased oxygen stress. The precise expression and function of PGC-1α in human brain are therefore of particular interest. To this end, several single-nucleotide polymorphisms (SNPs) at the PPARGC1A locus have been associated with HD and PD age of onset (19–22). In addition, PGC-1α expression was shown to decrease with dementia in AD brain (23), and a functional role in PD was recently suggested by studies using laser capture dissected analysis of dopamine-deficient neurons (24). Understandably, the lack of appropriate tissue/neuronal models makes human in vivo studies difficult, and most studies have relied on post-mortem analyses of expression differences between diseased and non-diseased tissues.

Many groups have emphasized the role of alternative splicing in the exquisite temporal and spatial control of transcription in eukaryotic species (25). Transcriptome studies predict that 92–94% of human genes encode multiple transcripts (26). Alternative promoter usage accounts for many of these transcripts (27), but internal exon incorporation or 3′ UTR alterations also add to transcriptional diversity (25). The PGC-1 family of transcriptional coregulators, including PGC-1α, PGC-1β and PGC-related protein (PPRC1), is no exception. Four distinct mRNA isoforms of human PGC-1β were identified that originated from a new promoter in intron 2 of the PGC-1 gene (30). Two muscle-specific isoforms (PGC-1α-b and PGC-1α-c) have also been described [UniProt.org (Q5VV67 and Q5VV67-2)] (29).

Regarding PGC-1α, a novel liver-specific isoform was identified that originated from a new promoter in intron 2 of the human gene (30). Two muscle-specific isoforms (PGC-1α-b and PGC-1α-c), initiating 6 kb and 13.7 kb upstream of exon 2, were selectively upregulated in response to exercise in rats (31). In pig, two exon 8 splice variants were described that differ in their distribution throughout various tissues (32). Recently, NT-PGC-1α utilizing an alternative 3′ exon distal to exon 6 that contains an in-frame stop codon exhibited different biological and functional properties compared with full-length PGC-1α (33). This isoform was also initiated from the NT exons found in muscle (34) and was shown to be altered in HD patients (35). However, new PGC-1α isoforms have yet to be described in brain.

Prompted by data implicating PPARGC1A in several neurodegenerative diseases and the increasing number of functional isoforms described at the PPARGC1A locus, we sought to identify potential new PPARGC1A isoforms in human brain that may give further insight into the pathophysiological role of the locus in neurodegenerative disorders. We report here the identification and initial characterization of several new brain-specific PGC-1α isoforms that arise from a common promoter almost 600 kb upstream of the current locus. More importantly, we show that this novel promoter is active in neuronal cell lines, and haplotypes encompassing the novel promoter are more strongly associated with HD age of onset than SNPs or haplotypes previously described for the reference locus (GenBank NM_013261.3).

RESULTS
Identification of new PPARGC1A isoforms in human brain
To test for the presence of alternative transcripts that may include a conserved region of PPARGC1A intron 2 near SNP rs7665116, shown to be significantly associated with HD age of onset in some populations (7,20,21), 5′-RLM-RACE was performed using outer primers specific to exon 4 and nested primers in exon 3. While a novel liver-specific isoform was detected using liver cDNA (30), no transcripts initiating in intron 2 were detected in the RLM-anchored brain cDNA that was prepared from total brain RNA of single individuals. Upon further evaluation of 5′-RLM-RACE products obtained using exon 2-specific primers, we discovered transcripts that initiated at the exon 1 as well as transcripts that contained several new exons that spliced to exon 2 (Fig. 1A). Further 5′-RLM-RACE experiments led to the identification of two additional 5′ exons (Fig. 1B).

Upon searching the UCSC database (36) [http://genome.ucsc.edu, February 2009 (GRCh37/hg19 assembly), the most 5′ positioned exon was located 587 kb upstream of exon 2 (Fig. 2A). Long-range polymerase chain reaction (PCR) using primers specific to exon B1 and exon 13 in brain cDNAs prepared from several individuals confirmed that these exons were not part of a pseudogene, but in fact represented bona fide transcripts. No amplicons were obtained in liver cDNA (Fig. 2B). In brain cDNA, a single wide band of ~2.5 kb was obtained that appeared more abundant than a product of similar size obtained with the same exon 13 primer and a control primer in exon 1 (Fig. 2B). Using the latter primer pair, a strong band was amplified from liver cDNA as expected.

Upon restriction digestion of several clones harboring amplicons generated from brain cDNA with the exon B1 and exon 13 primers, we found that the ~2.5 kb band represented at least five distinct transcripts that differed not only at the 5′ end, but also internally. Additional transcripts harboring different brain-specific exon combination were also detected (data not shown). Representative transcripts, depicted in Figure 2C, span a genomic region of almost 680 kb and appear to originate from a common promoter that has a highly predicted transcription start site. Furthermore, splice donor and acceptor sites of all new exons were highly conserved across species and adhered completely to the guanine thymine (GT)—adenine guanine rule (Supplementary Material, Figs. S1–S4). By performing a search of non-human spliced mRNAs in the UCSC database, we detected a transcript in Spalax galili (blind subterranean mole rat) that aligned with our transcript B1B4 exons 2–13 (JO012605).

Further characterization of new PPARGC1A transcripts in human brain
Northern blotting of brain polyA+ RNA with probes specific to exon B2/B5 showed two to three distinct bands (not detected in HepG2 polyA+ RNA) that migrated between ~6.5 and 7.5 kb (Fig. 3A, left panel). These RNA species correspond to full-length brain transcripts containing not only exons B2 and B5, but also exon B1 which is spliced to exon B5 and would result in a transcript ~400 nt larger than transcripts containing exon B2. The larger novel transcripts were also detected with
probes spanning exons 2–7 and exons 9–12 of the PPARGC1A reference locus (Fig. 3A, middle and right panels). Additional RNA species were detected at approximately 2.8 and 0.8–1 kb that most likely represent transcripts that contain exons B2 and/or B5 or exon 1 and, at least, part of exons 2–7, but not exons 9–12 of PPARGC1A (Fig. 3A, right panel).

Quantitative reverse transcription PCR (RT–PCR) using total brain cDNAs prepared from four individuals and cDNAs from a representative human tissue panel with primers spanning exons B1 and B4 as well as exons B5 and exon 2 or exon 1 and exon 2 showed that the new transcripts were mostly brain specific (Fig 3B). Interestingly, the relative abundance of brain transcripts differed in the four individuals and appeared to be more abundant than exon 1- and exon 2-containing transcripts. These data were further corroborated by RNase protection assays (RPAs) using cRNA probes spanning exons 1 and 2 (Fig. 3C). Protected fragments spanning the full exon 1/2 junction were the main transcripts detected in HepG2, liver and skeletal muscle, but in brain, additional protected fragments containing exon 2 were more abundant than transcripts spanning exons 1 and 2. These data indicated that exons other than exon 1 were spliced to exon 2 in brain. Faint bands of approximately 126 and 143 nt were also detected in liver and skeletal muscle that may represent low abundant exon 1-containing transcripts that are spliced to exons other than exon 2 and perhaps transcripts that are initiated upstream of exon 2.

Initial homology analyses of the new brain-specific exons using the UCSC database showed a high degree of conservation among higher vertebrates (Supplementary Material, Figs. S1–S4). We therefore used mouse brain to identify the transcripts harboring the brain-specific exons in the main cell types. Transcripts containing exons B1 and B4 were clearly identified in neuronal cells and oligodendrocytes and, to a lesser extent, in microglia, but not in astrocytes. Exon 1-containing transcripts were clearly amplified from astrocytes and, to a lesser extent, from neuronal cells and oligodendrocytes, but not from microglia (Supplementary Material, Fig. S5).

The new brain transcripts encode several PGC-1α isoforms

Detailed searches of the UCSC and ExPASY (37) (http://www.expasy.org) databases with the new transcript sequences showed that exons B4 and B5 contained methionine start codons in a Kozak consensus with open reading frames (ORFs) that continued in-frame into exon 2 of PGC-1α. Exon B4 appears to have evolved from a short interspersed element and the start codon and short ORF is fully conserved among higher vertebrates including the mouse (Supplementary Material, Fig. S2B). Exon B5 is also highly conserved among vertebrates and an ORF is predicted in many species except mouse (Supplementary Material, Fig. S3A). Exons B1, B2 and B3 are conserved at the sequence level (Supplementary Material, Figs. S1 and S2) and contain ORFs, but not in the context of the transcripts that we identified. Exon B1 is spliced to either exon B4 or B5 (Figs. 1B and 2C). The

Figure 1. Identification of new PPARGC1A exons in brain. (A) New brain exons B5, B3 and B2 obtained by 5′-RLM-RACE from exon 3 and by nested PCR from exon 2. (B) Further 5′-RLM-RACE and nested PCR with a downstream primer in exon B5 showing the existence of two additional 5′ exons that were termed B1 and B4.
ORFs with start codons in exons B1, B2 and B3 (Supplementary Material, Figs. S1A and B and S2A) are terminated early by stop codons in exon B4 or B5. Hence, exon B1-initiated transcripts would avail of the initiating start codons in exon B4 or B5. However, the possibility that exon B1, B2 or B3 are spliced to other exons, not identified, cannot be excluded. Full-length exon B2- and exon B3-containing transcripts spliced directly to exon 2 were minor populations as judged by long-range PCR (data not shown). Such transcripts do not have a predicted start codon in exon B2 or B3 and would use a downstream methionine in exon 3 as was shown for a recently described liver-specific isoform (30).

The new brain transcripts differed not only at the 5′ ends but also within the PGC-1α internal sequence that would predict new stop codons. A new splice variant containing a 204 bp exon 8a was found in association with exon B4 that was also detected in association with exon 1 in a chondrocyte library (http://www.ncbi.nlm.nih.gov/genbank/ AK309261, AK296591) and with two new 5′ exons derived from intron 2 in testis (AK301883). The predicted amino acid sequence of this isoform adds a new nine-amino acid extension after exon 7 (Fig. 4A and Supplementary Material, Fig. S3B) and would result in proteins of approximately 32–34 kDa depending on the start codons utilized (Fig. 4B). Interestingly, exon B5 was also detected in transcripts containing exon 7a that would predict an isoform very similar to NT-PGC-1α, originating from exon 1 (33). Exon B4 was also detected in transcripts containing a new 82 bp exon 3 extended (3ext) that predicts a VRTLPTV C-terminal peptide replacing the second LXXLL motif of PGC-1α encoded by the reference transcript (Fig. 4A and B and Supplementary Material, Fig. S3B).
The main clones (with an added His-tag) were translated in vitro using rabbit reticulocyte lysates and gave rise to proteins of the expected or slightly larger sizes (Fig. 4C). To determine if the predicted proteins exist in a neuronal cell model, western blot analyses were performed in SH-SY5Y cytoplasmic and nuclear extracts. A polyclonal NT antibody directed against the first 200 amino acids, that would likely detect both the PGC-1α and the brain-specific proteins, showed bands corresponding to the sizes predicted for the various isoforms (Fig. 4D), while a polyclonal C-terminal antibody detected the full-length protein (data not shown).

Localization of new PGC-1α isoforms

The subcellular localization of PGC-1α isoforms was determined by transfecting SH-SY5Y cells with the main new transcripts cloned in-frame into enhanced green fluorescent protein (eGFP) expression vectors and then visualizing GFP expression by confocal microscopy. As previously shown (1,30), full-length PGC-1α was localized to the nucleus (Fig. 5, left panel). Full-length B5-PGC-1α also localized to the nucleus (middle panel) and, like PGC-1α, appeared to be in nuclear sub-compartments or ‘speckles’ as previously noted by others. Full-length B4-PGC-1α also localized to the nucleus (data not shown), while B4-NT-PGC-1α was readily detected in the cytoplasm of SH-SY5Y cells exhibiting a localization pattern comparable to that of mouse NT-PGC-1α (33). B5-8a-PGC-1α appeared to have a universal distribution both in the cytoplasm and in the nucleus (data not shown). B4-NT-PGC-1α and B5-8a-PGC-1α lack the consensus nuclear localization signals mapped down-stream of their respective stop codons (7,38,39), but their predicted size of <35 kDa should not impede their diffusion through the nuclear pore complex (40). Like PGC-1α, the isoforms contain two consensus nuclear export signals (NES) near their NT.
exportin, CRM1, through its proximal NES have been shown to account for its predominant cytoplasmic localization (41). Nuclear export of NT-PGC-1α is inhibited by protein kinase A-mediated phosphorylation of specific residues (33). Whether similar mechanism(s) regulate the subcellular distribution of B4-NT-PGC-1α and B5-8a-PGC-1α remains to be determined.

Identification and characterization of the major brain PPARGC1A promoter

In order to map the most distal 5′ boundaries of the new brain transcripts, additional 5′-RLM-RACE was performed from exons B2 and B1. No additional sequence (notably that of exon B1) was found upstream of exon B2. We therefore assumed that exon B2-containing transcripts originate from a promoter within the 1 kb region between exons B1 and B2. In contrast, 5′-RLM-RACE from exon B1 resulted in three additional clones with 5′ extensions of 56, 43 and 38 bp (Supplementary Material, Fig. S6). This region is very guanine cytosine (GC) rich and may have hindered the RT-reaction in the original RACE protocol that led to the identification of exon B1. The most distal 5′ sequence of exon B1 coincided with a highly predicted transcription start site (SwitchGear Genomics, Gene Model CHR4_M0162) and lies within a CpG island as well as a neuron-specific brain histone H3K4me3 peak (Supplementary Material, Fig. S7) identified by chromatin immunoprecipitation sequencing of neuronal nuclei collected from the prefrontal cortex of 11 humans ranging in age from 0.5 to 69 years (42).

To identify and localize antisense transcripts and miRNA at the extended PPARGC1A locus, we searched the Natural Antisense Transcription database (http://natsdb.cbi.pku.edu.cn/, release July 2006), but found no antisense entries in the region. However, a search of the UCSC database for human spliced expressed sequence tags (ESTs) in the PPARGC1A antisense direction revealed four relevant entries. Two transcripts are located downstream of PPARGC1A and two other transcripts are located within 40 kb of the B1 transcription start site. These transcripts are shown in Supplementary Material, Figure S8, and their potential relevance is described in the figure legend. Finally, two miRNA were found near the new

---

**Figure 4.** Identified and putative PGC-1α isoforms in brain and SH-SY5Y cells. (A) B4, B5 and exon 1 contain independent translation start sites resulting in NT sequence differences shown on the left. B1, B2 and B3 are not translated in the context of transcript structures identified. Transcripts structures resulting from alternative splicing and their predicted amino-acid sequences are shown on the right. Bold lettering indicates sequence differences among predicted proteins. (B) Predicted amino acid content and expected size of new brain isoforms including those terminating with in-frame stop codons in exons 8a, 7a and 3ext. Isoforms shown in bold were present in the clones sequenced. (C) Western blot of *in vitro* translation reactions of the transcripts indicated (see Materials and methods). (D) Immunoblots of cytoplasmic (cyto) and nuclear (nuc) extracts from SH-SY5Y cells using an NT polyclonal antibody. Expected size ranges of the new isoforms are shown on the right in comparison with PGC-1α (Ex1-PGC-1α). X denotes possible differences in NT sequences due to differences in 5′ exons (see B).

**Figure 5.** Subcellular localization of PGC-1α, B5-PGC-1α and B4-NT-PGC-1α in SH-SY5Y cells. Fluorescence micrographs of SH-SY5Y cells transiently transfected with eGFP in-frame fusion constructs pPGC-1α-GFP, pB5-PGC-1α-GFP and pB4-NT-PGC-1α-GFP are shown. DAPI of MITO denotes nuclear staining with 4′,6′-diamidino-2-phenylindole or mitochondrial staining with MitoTracker RedCMXRos, respectively.
promoter region using Ensembl and UCSC databases. MIR573 (ENST00000384964) and a second miRNA (ENST00000410330) are located upstream of exon B1 or between exons B2 and B3, respectively (Supplementary Material, Fig. S8). A sequence search of MIRBase (http://www.mirbase.org, release 18, November 2011) showed that the latter miRNA is similar to the human miRNA-1302 family.

Using primers specific to the predicted promoter region and genomic DNA from control individuals, we cloned a ∼2.0 kb fragment upstream of exon B1 into the pGL4.11 luciferase vector and noted, in transient transfection assays, that the brain-specific promoter was at least equal or more active than a reference gene promoter construct of similar length in SH-SY5Y and NTERA-2D cell lines (data not shown). Deletions from ∼2 to 1.2 kb to 556 bp increased the transcriptional activity of the new promoter. The 2 kb promoter was cloned and sequenced from three control individuals and revealed several conserved transcription factor binding sites, notably an USF1/Myc/Max/SREBP-1c site within the 556 bp fragment. Co-transfections of USF1 expression plasmids enhanced the transcriptional activity of the 556 bp promoter as expected (Fig. 6).

Relevance of the new brain promoter for HD

The European HD network cohort is described in Supplementary Material, Table S1. HD cytosine-adenine-guanine (CAG) and non-HD CAG repeat sizes and their product explained 62% of the variability in the population. Previous studies have shown that some SNPs in the promoter region and intron 2 of the PPARGC1A reference gene are associated with HD age of onset, suggesting that PPARGC1A is a modifier gene of HD. As the new brain promoter is located in a haplotype block that is far upstream of the haplotype blocks comprising the PPARGC1A reference locus (Fig. 7), we ascertained whether the new region also shows such associations. We typed SNPs rs17592631, rs2048025 and rs11737023 in the well-characterized HD cohort of the European HD network. In addition, we sequencing studies identified two microsatellite regions located at −150 and −10 bp relative to the transcription start site. While the more proximal dinucleotide repeat showed no variation in 50 subjects analyzed, the more distal GT dinucleotide repeats near rs6448272 showed considerable variation in the repeat size. We sequenced PCR products in all 1706 subjects included in the study and manually read the allele sequences as well as rs6448272. The GT insertion polymorphism ranged from 9 to 28 repeats. rs6448272 major and minor alleles were in perfect linkage disequilibrium (LD) with <14 and >13 GT repeats, respectively (Supplementary Material, Table S2). A highly significant association was noted between GT repeats of the two alleles (R = 0.4649, P < 0.0001). Intriguingly, the sum of GT repeats in both alleles or the GT repeat size of the allele containing more repeats (upper allele) were associated with the CAG repeat size of the HD allele (Pearson’s R = 0.0667, P = 0.0044 or R = 0.0657, P = 0.0067). These results were confirmed using the non-parametric Spearman test (R = 0.0663, P = 0.0061 or R = 0.0587, P = 0.0152).

The correlation of the GT repeat size of the allele containing fewer repeats (lower allele) with the CAG repeat number of the HD allele did not reach significance in parametric (P = 0.0648) or non-parametric testing (P = 0.0852). None of the SNPs alone or the GT repeat size of either allele showed an association with HD age of onset (Supplementary Material, Table S3). For haplotype analysis, the GT insertion polymorphism was separated into three loci with genotypes of <14 or 14–17 repeats (GT I), <18 or 18–21 repeats (GT II) and <22 or >21 GT repeats (GT III). As rs6448272 showed an R² of 1.00 with GT I (Supplementary Material, Table S2), it was omitted from the analysis. A significant global haplotype effect was noted in the model adjusted for HD CAG and non-HD CAG repeat sizes and their product. Relative to the most common haplotype 111111, haplotypes 122111 and 111112 were found to be protective (Table 1). This result was substantiated with another multivariate model using an expectation substitution method for haplotypes (43) as well as haplotype score testing (44) (Supplementary Material, Table S4). We next compared the effects of the new promoter haplotypes with PPARGC1A SNPs that showed associations with HD age of onset in populations studied previously. We noted a borderline significant association of rs2970870, but no association of rs76651166 in the current study population (Supplementary Material, Table S3).

DISCUSSION

Alternative splicing and/or transcript initiation substantially increase the complexity of mammalian transcriptomes. In humans, the majority of protein-coding transcriptional units have one or more alternative promoters (27). The data presented here show that the human PPARGC1A locus is ~6-fold larger than originally described (3). It contains a novel promoter that is located in a large CpG island 583 kb upstream of exon 1. From the new promoter, several brain-specific transcripts are initiated that are likely more abundant than the transcripts originating from the reference gene promoter. Furthermore, EST data and our own RACE experiments (data not shown) imply that transcripts from other tissues such as kidney and ovary may also be initiated from within this ~580 kb upstream region. As exemplified by the association of haplotypes in the extended region of the PPARGC1 locus with HD age of onset, our findings in
human brain tissue are likely relevant for neurodegenerative disorders.

The novel PPARGC1A promoter and the brain-specific transcript structures were deduced by complementary methods including 5'-RLM-RACE, northern blots, predictions from promoter algorithms and transient transactivation assays. Unlike in human liver, where a recently described transcript, initiated in intron 2, is spliced to exon 3 (30), all the variant brain-specific transcripts identified are spliced to exon 2. Thus, the full-length brain-specific transcripts contain the new exons identified and reference gene exons 2–13 in a regular order. Additional full-length transcripts contain interspersed sequences resulting in truncated proteins.

The brain-specific promoter region contains numerous putative binding sites for various transcription factors (Supplementary Material, Fig. S7). Among these sites is a highly conserved E-box that can be targeted by several transcription factors including USF1 and SREBP-1c. To this end, we demonstrated transcriptional activation of the novel promoter by USF1 in transient transfection studies of SH-SY5Y cells. Noteworthy, de novo cholesterol synthesis and nuclear levels of SREBPs are reduced in HD brain tissues (45,46). As a result, transcriptional activation of brain-specific PPARGC1A may be reduced. The presence of a CpG island and a brain-specific chip signal for trimethylated H3K4, commonly associated with transcriptional regulation (42,47), in the brain-specific promoter region suggests a role for epigenetic regulation. The promoter also contains an instable microsatellite region consisting of GC repeats followed by a variable number of GT repeats. This gene segment containing between 60 and 78 bp showed a high score for Z-DNA formation (Z-score $\approx 500,000$) as calculated by the ZHunt (48) (http://gac-web.cgrb.oregonstate.edu/zDNA/). Potential Z-DNA sequences are non-randomly distributed in the genome, occur more often proximal to transcription start sites of transcribed genes (49) and have been shown to modulate transcription in a context-dependent manner (50). To our knowledge, HIF-1α is the only transcription factor that has been shown so far to bind to such regions (51). Within the brain-specific promoter lies a predicted, but not further characterized, gene locus (hypothetical protein LOC729175, NCBI Ref. sequence XP_001129558.1) that contains, between predicted amino acids 80 and 253, a domain exhibiting homology to DNA polymerase gamma and tau. The hypothetical protein is transcribed from the same strand and in the same direction as the brain-specific PPARGC1A transcripts. The EST clones of this locus have been described in ES cells (GenBank Nr. CN283176) and overlap the region harboring exons B1 and B2 (Supplementary Material, Fig. S7). Whether this locus is relevant for genomic instability awaits further study.

In comparison with transcripts initiated at the reference gene promoter, the brain-specific transcripts encode distinct NTs. PGC-1α encoded by the reference transcript has a strong activation domain at the NT that interacts with histone acetyltransferase complexes such as SRC-1 and CREB-binding protein (52). Whether the functions of these domains are altered in brain-specific isoforms requires

Table 1. Brain-specific promoter haplotypes and HD age of onset

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Frequency</th>
<th>Age of onset (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>111111</td>
<td>0.391</td>
<td>1.314 (1.233–1.397)</td>
<td>Reference</td>
</tr>
<tr>
<td>122111</td>
<td>0.305</td>
<td>1.321 (1.239–1.404)</td>
<td>0.0398</td>
</tr>
<tr>
<td>211112</td>
<td>0.077</td>
<td>1.322 (1.239–1.405)</td>
<td>0.2249</td>
</tr>
<tr>
<td>111112</td>
<td>0.061</td>
<td>1.337 (1.254–1.420)</td>
<td>0.0006</td>
</tr>
<tr>
<td>121112</td>
<td>0.047</td>
<td>1.320 (1.237–1.403)</td>
<td>0.3870</td>
</tr>
<tr>
<td>122211</td>
<td>0.046</td>
<td>1.310 (1.226–1.394)</td>
<td>0.4786</td>
</tr>
<tr>
<td>122112</td>
<td>0.030</td>
<td>1.303 (1.218–1.387)</td>
<td>0.1916</td>
</tr>
<tr>
<td>211111</td>
<td>0.019</td>
<td>1.315 (1.230–1.387)</td>
<td>0.9988</td>
</tr>
</tbody>
</table>

Expected phenotypic means for age of onset are logarithmically transformed values. $P = 0.0112$ for global haplotype effects.

$^a$1 or 2 refers to the major or minor alleles, respectively, in the following order: rs17592631 C/T, rs2048025 C/T, GT repeat locus I $<14/>13$, GT repeat locus II $<18/>17$, GT repeat locus III $<22/>21$, rs11737023 G/A.

Figure 7. Haplotype blocks at the PPARGC1A locus (HapMap Data Phase III/Rel#2, February 2009, on NCBI B36 assembly, dbSNP b126, CEU population). Exons identified in brain transcripts are shown at the top. The PPARGC1A locus as previously described along with DHX15 [DEAH (Asp-Glu-Ala-His) box polypeptide 15] are shown in the middle. The SNPs and GT insertion polymorphism typed in this study are shown in relation to haplotype blocks.
further investigation. Apart from differences at the NT, various splice variants generated downstream of exon 3 add to the complexity of the PPARGC1A locus. In addition to the full-length proteins and an NT-PPARC-1α-like protein, two additional isoforms and their respective transcripts have been observed. One isoform showed an extension of the exon 7-encoded sequence by nine amino acids. The other protein isoform is characterized by a seven amino acid extension of the sequence encoded by exon 3. The latter isoform may be of particular relevance, as the new amino acid segment replaces the second LXXLL box of PGC-1α known to be essential for the coactivation of numerous nuclear receptors. It is therefore conceivable that this isoform has a dominant negative effect. Alternatively, this isoform may be recruited to different targets. Thus, the new transcripts that we describe may provide a mechanism for separation of mitochondrial and nuclear function as these transcripts encode proteins that differ not only at their NTs but also in internal domains that alter their subcellular localization.

The new brain promoter lies in a distinct haplotype block clearly separated from the regions encoding the reference gene promoter. The association of haplotypes with HD age of onset in the well-characterized large HD cohort of the European HD network strongly suggests that the expression of brain isoforms plays a role in the pathogenesis of HD. Unexpectedly, rs7665116 previously found to be associated with age of onset in an Italian HD cohort showed no associations in the larger European HD cohort. These contrasting results may be explained by the genetic backgrounds of the different populations. Indeed, the minor allele frequency in the current population was lower than that in the previous population from Italy. It is also possible that the collection of subjects from different European regions and in different centers increased the signal-to-noise ratio in the current study. Thus, the association between the brain-specific promoter haplotypes with HD age of onset is probably robust and its true strength may have been underestimated. An intriguing finding was the association of the GT repeat length with rs6448272 located in the transcribed region of the hypothetical protein LOC729175. Whether rs6448272, which predicts a synonymous change in an arginine codon, or another SNP in LD with it, influences the GT repeat size, remains to be determined. Another potentially interesting result was the correlation between the GT repeat size of the allele containing more repeats and the CAG repeat size or the HD allele. Currently, we have no clear explanation for this finding.

In most studies to date (human and mouse) that have measured PPARGC1A mRNA expression, primers were used in exon 3–5 regions which now in hindsight would measure not only PPARGC1A reference sequence transcripts but also the majority of brain transcripts encoding full-length and truncated isoforms. Considering differences in regulation of brain-specific and reference sequence transcripts as well as potential functional differences between full-length and truncated isoforms, it will be of utmost importance to distinguish which PGC-1α isoforms are affected in distinct human disorders. Enhanced alternative splicing driven by RNA polymerase III-transcribed RNA has been implicated in neurodegeneration (55), and aberrant splicing in motor neurons has been associated with sporadic ALS (54). The potential new PGC-1α protein repertoire in the human brain requires careful analysis before further attempts are made to selectively increase the expression of the PGC-1α in neurodegenerative disease models. Indeed, sustained over-expression of PGC-1α impaired dopaminergic function in vivo, highlighting the need for maintaining a physiological level of its activity (55). Newer treatments may be fine-tuned to target the brain-specific isoforms in order to result in more physiological effects.

**MATERIALS AND METHODS**

**Clinical resource**

For the modifier study, the human genetic material and clinical information were obtained in 2009 through the European Huntington’s Disease Network (EHDN) REGISTRY. The EHDN REGISTRY project is a multinational observational study; more details are available in reference (56) or at http://www.euro-hd.net/html/registry. The participating centers from 16 European countries are listed in the Supplementary Material, Appendix. The data recorded from participants include the result of the CAG repeat length reported by the local service laboratory, age of onset, gender and sibship information. Participants have an option of donating fresh blood samples, which are taken in acid citrate dextrose (ACD) tubes Vacutainer, Becton Dickinson, Milan, Italy) and couriered to the central laboratory, BioRep, Milan. This material was used for the SNP genotyping.

In total, we included 1706 patients from 16 countries (Supplementary Material, Table S1) with genetically confirmed HD, known length of the expanded and unexpanded CAG repeat allele and known age of onset. From known sibships, only one individual was included. Age of onset was defined as the age at which first HD symptoms appeared as judged by a trained neurologist either from the neurological examination or (more frequently) from the patient history as recorded in the REGISTRY. When different ages of onset for motor or psychiatric/cognitive symptoms were given, the earlier age of onset was used.

**Genotyping**

We typed rs2970870 and rs7665116, previously shown to be associated with HD age of onset, using TaqMan Genotyping Assays (Applied Biosystems) C_1643241_10 and C_31279675_10, respectively. SNPs located in the promoter region of the newly described brain transcripts included rs2048025, rs11737023 and rs17592631. The respective typing reagents were C_473389_10, C_1222369_10 and C_32808047_10. rs6448272 genotypes as well as the number of GT insertions at ~150 bp relative to the novel transcription start site were determined by sequencing using a DNA fragment obtained by PCR using primers described in Supplementary Material, Table S5. The accuracy of the GT number determination was verified by sequencing the second PCR products in 30 patients with GT repeat sizes ranging from 10 to 28 and cloning the representative alleles into the KpnI/HindIII sites of pGL4.11 (Promega). In all cases, initial and repeat analyses showed identical results.
RNA sources

Total brain RNA from four individuals and a human multiple tissue panel including a human brain pool were purchased from Ambion. Brain and kidney polyA+ RNA were purchased from Clontech. Total RNA from human muscle biopsies, mouse neurons, astrocytes, microglia and oligodendrocytes as well as from HepG2 cells and SH-SY5Y and differentiated NTERA-2D cells was prepared with a Qiagen RNeasy Lipid Tissue Midi kit (Qiagen). PolyA+ RNA was extracted from approximately 1 mg total muscle or HepG2 RNA using a PolyAtract® mRNA Isolation System (Promega).

5′-RLM-RACE and transcript cloning

Total brain RNA was treated and amplified with FirstChoice® RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE) kit (Ambion) according to the manufacturer’s protocol. Capped cDNAs were amplified by touchdown PCR using external and nested primers corresponding to the 5′-RACE adaptor sequence or complementary to PPARGC1A (listed in Supplementary Material, Table S5B). RACE products were subcloned into the PGEM®-T-Easy vector (Promega) and sequenced using a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and the ABI 3500 genetic analyzer (Applied Biosystems). Exon-specific primers were subsequently designed from the new RACE products and subjected to long-range PCR using the expand PCR template kit (Roche) and primers specific to PPARGC1A (Supplementary Material, Table S5).

Plasmids

Northern blot and RPA probes were made by cloning the respective PCR amplification products into the pGEM®-T-Easy vector followed by linearization with the appropriate restriction enzyme to generate cRNA transcripts. Plasmids used for in vitro transcription/translation included the coding sequences of the wild type, B4-PGC-1α, B5-PGC-1α, B4-7a-PGC-1α and B4-8a-PGC-1α amplified using primers listed in Supplementary Material, Table S5, and cloned into the BamHI/XhoI sites of pT7CFe1-Chis vector (Thermo Scientific). For isoform localization studies, a similar strategy was used to clone the same coding sequences into the XhoI/BamHI sites of the pEGFP-N1 vector (Clontech) with primers listed in Supplementary Material, Table S5. Plasmids used for promoter luciferase assays included a 2.1 kb segment (#1936 to +58) with +1 defined as the transcription start site of exon B1) of the brain PGC-1α promoter amplified and cloned into the KpnI/HindIII sites of pGL4.11 [Luc2P] (Promega). Deletion constructs were produced by Swal and XhoI digestion of the full-length vector and re-ligation to produce −1.2 kb and −556 bp fragments, respectively. pcDNA6/v5HisA expression plasmids (USF1 ad PGC-1α) have been described previously (30).

Quantitative RT–PCR

DNase I-treated total brain RNA (1 μg/reaction) from four separate individuals and from a human tissue panel was reverse transcribed using random hexamers and/or a poly-15dT primer and moloney murine leukemia virus RT kit (Invitrogen). cDNAs were amplified using iQ™ SYBR Green Supermix (Bio-Rad) and primers listed in Supplementary Material, Table S5. Constitutively expressed RPLP0 (Ribosomal Protein, large, P0) RNA was used for normalization of mRNA abundance. Relative mRNA levels were calculated using the comparative threshold cycle method (ΔCt) and the iCycler iQ Multicolour Real-Time PCR Detector along with the GeneX software (Bio-Rad).

Northern blot analyses for brain-specific transcripts

A NorthernMax® kit (Ambion) was used for northern analyses with 5 μg of human brain (Ambion), 2.5 μg muscle and kidney polyA+ RNA (Clontech), or 5 μg HepG2 poly(A)+ RNA per lane separated in 1.1% denaturing agarose gels. RNA was transferred to BrightStar®-Plus Positively Charged Nylon Membrane (Ambion) using a Turboblotter™ system (Whatman/Schleicher and Schuell). Membranes were hybridized in UltraHyb® northern blot solution (Ambion) with P32-CTP-labeled RNA probes complementary to the new brain or reference sequence PPARGC1A coding regions. Blots were washed at 68°C with low and high stringency buffers and subsequently exposed to Amersham Hyperfilm™ MP (GE Healthcare).

RNase protection assays

RPAs were performed with the RPA III™ Ribonuclease Protection Assay Kit (Applied Biosystems/Ambion) as described (30) except with DNA plasmid templates for in vitro transcription of P32-CTP-labeled antisense RNA probes spanning exons 1 and 2 resulting in specific PPARGC1A transcript sequences (Supplementary Material, Table S5).

In vitro transcription/translation and western blot analysis

TNT® Quick Coupled Transcription/Translation and Transcend™ Chemiluminescent Non-Radioactive Translation Detection Systems (Promega) were used for in vitro synthesis of PGC-1α proteins. Circular plasmids (approximately 1 μg) pCFe-PGC-1α, pCFe-B4-PGC-1α, pCFe-B5-PGC-1α, pCFe-B4-7a-PGC-1α and pCFe-B4-7b-PGC-1α were each incubated with 25 μl rabbit reticulocyte lysate mix, 1 μl of non-radioactive amino acids and 1 μl Transcend™ precharged e-labeled biotinylated lysine-tRNA complex at 30°C for 90 min. Samples were denatured in sample buffer (62.5 mM Tris–HCl, pH 6.8, 2% w/v SDS, 10% glycerol, 50 mM dithiothreitol, 0.01% w/v bromphenol blue) at 70°C for 10 min, cooled on ice and subjected to electrophoresis in 8% SDS–polyacrylamide gels. Gels were transferred to polyvinylidene fluoride (PVDF) membranes (Amersham) that were blocked with 1 x tris-buffered saline (TBS), 0.1% Tween 20 for 1 h, incubated for 45 min with streptavidin-horseradish peroxidase (HRP) conjugate (1:7500 diluted), washed, incubated with 5 ml substrate mix for 1 min and exposed to the Image Station 20000 Multi-Modal Imaging System (Eastman Kodak Co.).

Brain whole tissue and nuclear lysates from human adult normal tissues were purchased from ABCAM. SH-SY5Y
and NTERA-2D nuclear and cytoplasmic extracts were prepared using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Pierce) according to the manufacturer’s protocol. Lysates (5–50 μg) were denatured for 95°C for 5 min in SDS sample buffer, separated in 7–10% SDS-polyacrylamide gels and transferred to PVDF membrane. Membranes were blocked as before and then incubated overnight at 4°C with NT or C-terminal [PGC-1 (P-19); sc-5815 or PGC-1 (K-15); sc-5816, respectively, (Santa Cruz)] antibodies diluted 1:250 in 1 × TBS/0.1% Tween 20 (TBST)/5% BSA. Blots were washed several times with 1 × TBST and then incubated for 1 h at room temperature with secondary antibody [Donkey anti-goat IgG–HRP: sc-2020 (Santa Cruz)] diluted 1:10,000 in 1 × TBST. After extensive washing with 1 × TBST, blots were incubated with SupernSignal® West Dura Extended Duration Substrate (Thermo Scientific) and imaged as before.

**Confocal microscopy**

SH-SY5Y and NTERA-2D cells were plated on cover slips in six-well plates and transfected for 24 h with 500 ng plasmid pB4-PGC-1α-GFP, pB5-PGC-1α-GFP, pB4-PGC-1α-7a-GFP, pB4-PGC-1α-7b-GFP or pPGC-1α-GFP. Cells were stained with 100 nM MitoTracker® Red (Invitrogen) for 45 min, rinsed three times for 15 min with 1 × phosphate buffered saline (PBS), fixed with 4% paraformaldehyde, rinsed with PBS, stained with DAPI (4′,6-diamidino-2-phenylindole, Sigma-Aldrich) for 30 min and rinsed finally in PBS. Slides were mounted with DAPCO and Mowiol. A Zeiss LSM710 confocal microscope equipped with an Axiocam digital camera and an oil-immersion ×63 objective lens was used for microscopy.

**Cell culture and transfection experiments**

SH-SY5Y and NTERA-2D1 cells were obtained from ATCC and cultured in DMEM/F12 1:1 (Invitrogen) and MEM (Invitrogen), respectively. Both media were supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin (Invitrogen). Cells were plated in 24-well dishes one day before transfection and transfected with plasmid constructs for 24 h using Lipofectamine™ 2000 (Invitrogen). Luciferase activities were measured using a Dual Luciferase® Reporter Assay System (Promega) as described. Results are representative of three experiments, each performed in quadruplicate, and are given as means ± SD.

**Isolation and culture of mouse brain cells**

C57Bl/6 mice (1–2 days old for neuronal and oligodendroglia cell cultures; 1–5 days old for microglia and astrocyte cell culture) were decapitated according to the guidelines of the Animal Research Center of the University of Ulm, Ulm, Germany.

Microglia were prepared as described (57). Briefly, neopallia were dissected and enzymatically (1% trypsin, Invitrogen, 0.05% DNase, Worthington, 2 min) dissociated. The resulting cells were centrifuged (200g, 10 min) and suspended in DMEM (Invitrogen) supplemented with penicillin (100 U/ml), streptomycin (100 μg/ml) and 10% heat inactivated fetal bovine serum (PAA Laboratories). Cells of two brains were plated into 75 cm² flasks (BD Falcon) pre-coated with 1 μg/ml poly-L-ornithine (Sigma). Cells were washed after 3 days. When abundant microglia were visible on an astrocyte monolayer (after 7–10 days in culture), microglia were manually shaken off and centrifuged (200g, 10 min). Pelleted cells were immediately frozen at −80°C. Astrocytes were prepared as microglia, except that cells corresponding to half a brain were plated in 10 ml culture medium per flask. After forming a confluent astrocyte monolayer (after 10 days), cells were washed three times with PBS and trypsinized with Trypsin/EDTA (Invitrogen). Cells were pelleted by centrifugation (200g, 10 min) and frozen at −80°C.

Oligodendrocytes were prepared as described (58). Briefly, the neopallia were dissected, enzymatically (0.15% papain, 0.04% L-cystein, 0.006% DNase, 20 min) and mechanically dissociated, centrifuged (300g, 5 min), suspended in the culture medium described above and plated into 75 cm² flasks (BD Falcon) pre-coated with 1 mg/ml poly-L-lysine (Sigma). Cells corresponding to two brains were plated in 10 ml culture medium per flask. Culture medium was changed 3–4 h later. After 3 and 6 days, two-thirds of the medium was changed and 5 μg/ml insulin was added. After 9 days in culture, microglia were removed by manually shaking the flasks and the oligodendrocyte precursor cells were removed manually by vigorous shaking the flasks. Cells were incubated at 37°C in a petri-dish for 30 min, pelleted by centrifugation (300g, 5 min) and immediately frozen at −80°C.

For the neuronal cultures, the neopallia were dissected and enzymatically (1% trypsin, 0.05% DNase, 10 min) and mechanically dissociated. The cell suspension was filtered using a cell strainer (100 μm) and centrifuged (200g, 5 min). Cells were resuspended in culture medium and 300 × 10⁶ cells were plated into a 75 cm² flask pre-coated with 1 μg/ml poly-L-lysine (Sigma). After 1 h, the medium was changed to Neurobasal medium supplemented with penicillin (100 U/ml), streptomycin (100 μg/ml), 2% B27, 2 mm L-glutamine and 10 μM glutamate. Half of the medium was changed after 1 and 4 days and 10 μM cytosine arabinofuranoside was added. After 7 days, the neurons were trypsinized, collected by centrifugation and frozen at −80°C.

**Statistical analysis**

Allele frequencies were estimated by gene counting. Agreement with Hardy–Weinberg equilibrium was ascertained using a χ² goodness-of-fit test. Correlations between the microsatellite GT content in the new brain promoter and the CAG content of the HD alleles were ascertained by the Pearson and Spearman rank correlation coefficient. Associations with age of onset were ascertained in linear models. As we observed a skewed relationship between the expanded CAG repeat size and age of onset as expected, we used logarithmically transformed age of onset as the dependent variable and individual SNPs, normal and expanded CAG repeat sizes as well as their interactions as independent variables (59). The GT insertion polymorphism was separated into three loci. Thus, at the first locus, GT repeat sizes <14 and >13 were coded as 1 and 2, respectively. At the second locus, GT repeat sizes <18 and >17 were coded with 1 and 2, and for the third locus, GT repeat sizes <22 and >21 were coded.
with 1 and 2. Assumptions of linear models were fulfilled, as linear relationships between variables and a normal distribution of residuals were observed.

The THESIAS software (http://genecanvas.ecgene.net/downloads) was used to estimate standardized pair-wise LD expressed in terms of D', haplotype frequencies and covariate-adjusted mean effects of haplotypes being present with a predicted frequency >0.015 on logarithmically transformed age of onset. As residuals are not obtained with the THESIAS software, we used an additional multivariate regression model to determine the distribution of residuals (43,60). For each subject, the expected number of copies of the haplotypes of interest carried by that individual was calculated at convergence of the EM algorithm (61). The expected haplotype frequencies were used as independent variables along with the covariate information. Using this expectation substitution method (in which unobserved true haplotype counts are replaced with their expected values given the haplotype frequency estimates and individual genotype data), we observed a linear relationship between HD CAG repeat size and logarithmically transformed age of onset. Furthermore, the residuals appeared to follow the normal distribution. Thus, the linear model appeared to provide an adequate fit to our data. As a third method for estimation of haplotype effects, we used the score test (44). This method is based on efficient score statistics, provides both global and haplotype-specific tests, allows the exclusion of rare haplotypes as well as adjustments for covariates. Furthermore, haplotype-specific scores for quantitative traits have been shown to fairly robust to departures from a normal distribution.

SUPPLEMENTARY MATERIAL
Supplementary Material is available at HMG online.

Conflict of Interest statement. None declared.

FUNDING
This study was supported by grants from the Fonds zur Förderung der wissenschaftlichen Forschung (FWF Project P19893-B05) to W.P., the European Huntington’s Disease Network (EHDN) REGISTRY to W.P. and P.W., the Thierry Latran Foundation (Pilot Project AAP09 15 01 73) to P.W. and the Verein für Medizinische Forschung Salzburg, Austria.

REFERENCES


