



HAL
open science

Upstream open reading frame-introducing variants in patients with primary familial brain calcification

Anne Rovelet-Lecrux, Antoine Bonnevalle, Olivier Quenez, Wandrille Delcroix, Kévin Cassinari, Anne-Claire Richard, Anne Boland, Jean-François Deleuze, Cyril Goizet, Alice Rucar, et al.

► **To cite this version:**

Anne Rovelet-Lecrux, Antoine Bonnevalle, Olivier Quenez, Wandrille Delcroix, Kévin Cassinari, et al.. Upstream open reading frame-introducing variants in patients with primary familial brain calcification. *European Journal of Human Genetics*, 2024, 32 (7), pp.779-785. 10.1038/s41431-024-01580-4. hal-04659594

HAL Id: hal-04659594

<https://normandie-univ.hal.science/hal-04659594>

Submitted on 23 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Upstream open reading frame-introducing variants in patients with primary familial brain
2 calcification.

3 Anne Rovelet-Lecrux^{1,*}, Antoine Bonnevalle^{2,*}, Olivier Quenez¹, Wandrille Delcroix¹, Kévin Cassinari¹,
4 Anne-Claire Richard¹, Anne Boland³, Jean-François Deleuze³, Cyril Goizet⁴, Alice Rucar⁵, Christophe
5 Verny⁵, Karine Nguyen⁶, Magalie Lecourtois^{1,*}, Gaël Nicolas^{1,*,#}

6 1 Univ Rouen Normandie, Inserm U1245 and CHU Rouen, Department of Genetics and CNRMAJ, F-
7 76000 Rouen, France

8 2 Univ Rouen Normandie, Inserm U1245 and CHU Rouen, Department of Neurology and CNRMAJ, F-
9 76000 Rouen, France

10 3 Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine (CNRGH), 91057,
11 Evry

12 4 Department of Medical Genetics, National Reference Center for Rare Diseases 'Neurogenetic',
13 Pellegrin Hospital, Bordeaux University Hospital, and University of Bordeaux, CNRS, INCIA, UMR 5287,
14 NRGGen Team, Bordeaux, France.

15 5 Department of Neurology, University-Hospital of Angers, 49933 Angers, France ; Unité MitoVasc,
16 UMR CNRS 6015, INSERM U1083, 49933 Angers, France

17 6 AP-HM, Hôpital Timone, Département de Génétique Médicale, Marseille, France

18 *equal contribution

19 Correspondence to: Gaël Nicolas, Inserm U245, UFR Santé, 22, boulevard Gambetta, 76000 Rouen,
20 France; 0033 2 35 14 82 80, gaelnicolas@hotmail.com

21

22

23 **ABSTRACT**

24 More than 50% of patients with primary familial brain calcification (PFBC), a rare neurological disorder,
25 remain genetically unexplained. While some causative genes are yet to be identified, variants in non-
26 coding regions of known genes may represent a source of missed diagnoses.

27 We hypothesized that 5'-Untranslated Region (UTR) variants introducing an AUG codon may initiate
28 mRNA translation and result in a loss of function in some of the PFBC genes.

29 After reannotation of exome sequencing data of 113 unrelated PFBC probands, we identified two
30 upstream AUG-introducing variants in the 5'UTR of *PDGFB*. One, NM_002608.4:c.-373C>G, segregated
31 with PFBC in the family. It was predicted to create an upstream open reading frame (ORF). The other
32 one, NM_002608.4:c.-318C>T, was found in a simplex case. It was predicted to result in an ORF
33 overlapping the natural ORF with a frameshift. In a GFP reporter assay, both variants were associated
34 with a dramatic decrease in GFP levels, and, after restoring the reading frame with the GFP sequence,
35 the c.-318C>T variant was associated with a strong initiation of translation as measured by western
36 blotting.

37 Overall, we found upstream AUG-introducing variants in the 5'UTR of *PDGFB* in 2/113 (1.7%)
38 undiagnosed PFBC cases. Such variants thus represent a source of putative pathogenic variants.

39

40

41 **Keywords**

42 Primary brain calcification, 5'UTR, upstream open reading frame

43

44

45

46

47 INTRODUCTION

48 Primary Familial Brain Calcification (PFBC) is a rare disorder characterized by microvascular calcification
49 in the basal ganglia and other cerebral regions. Patients with a pathogenic variant in one of the 6 genes
50 currently associated with PFBC exhibit brain calcifications on CT scan with full penetrance by the age
51 of 50 years. They may experience various neurological or psychiatric symptoms with incomplete
52 penetrance, such as movement disorders, gait disturbance, mood disorders, psychotic symptoms, or
53 cognitive impairment (1, 2). Four autosomal dominant genes (*SLC20A2*, *XPR1*, *PDGFB*, *PDGFRB*) and
54 two autosomal recessive genes (*MYORG*, *JAM2*) have been identified as PFBC-causative so far (3-9)
55 while some other genes are associated with differential diagnoses (10), including the recently reported
56 *CMPK2* gene (11). However, more than half of the probands remain genetically unexplained,
57 suggesting the existence of other genes as well as putative cryptic variants in known genes.

58 Pathogenic variants in the above-mentioned genes cause PFBC through a loss-of-function effect, either
59 through haploinsufficiency (deletions, premature stop codons) regarding *SLC20A2*, *PDGFB*, *MYORG*
60 and *JAM2* (3, 5, 7-9, 12, 13), or through missense variants impairing the protein levels and/or functions,
61 in all 6 genes (6, 14-17). Most of the time, variant interpretation in a medical setting is restricted to
62 coding exons and intronic boundaries. However, variants in non-coding regions, such as introns, 5'
63 untranslated regions (UTRs), 3'UTRs, or regulatory regions, may disturb protein levels by decreasing
64 transcription, altering splicing, increasing mRNA degradation, or impairing translation, thus causing
65 some degree of loss of function. 5'UTRs of mammalian genes are conserved regions that are
66 considered as important regulators of protein levels (18). They indeed repress gene expression through
67 diverse mechanisms. In addition, some genes contain natural upstream open reading frames (ORF) in
68 their 5'UTR, which further repress gene expression by initiating and terminating translation of rather
69 short peptides, upstream of the natural site, hence reducing the translation efficiency at the natural
70 translation initiation site (19). Variants in 5'UTRs may affect gene expression through diverse
71 mechanisms (20). They may alter natural upstream ORFs or introduce an AUG codon in a nucleotide
72 context compatible with translation initiation (18), leading to a novel upstream ORF. Such "mutant"
73 upstream ORF can be classified into three categories: (i) presence of a stop codon before the natural
74 translation initiation site (uORF), putatively repressing the translation initiation at the natural site or
75 leading to mRNA degradation, (ii) absence of a stop codon before the natural translation initiation site
76 while mutant ORF being in frame with the natural ORF, thus creating an overlapping in-frame ORF
77 leading to an N-terminally elongated protein (eORF) and (iii) absence of a stop codon before the natural
78 translation initiation site while mutant ORF being out of frame with the natural ORF, thus creating a
79 frameshift overlapping ORF (oORF).

80 Here, we reannotated exome sequencing data of 113 unrelated PFBC patients without a known
81 pathogenic variant in one of the six known genes and identified two variants predicted to introduce a
82 novel AUG in the 5'UTR of the *PDGFB* gene. We introduced these variants in a reporter assay and
83 showed that they both likely lead to decreased expression.

84

85 **METHODS**

86 *Patient selection and exome sequencing*

87 A multicenter national patient recruitment in France was performed followed by exome sequencing,
88 as previously described (1). Briefly, patients requested a genetic analysis in a clinical setting in the
89 context of suspected PFBC. They all signed a consent for genetic analysis in a medical setting. Criteria
90 used for genetic screening were the following: presence of brain calcification with a total calcification
91 score above the age-specific threshold (21) and normal parathormone, phosphate and calcium levels
92 in blood. Genetic screening was performed either by *SLC20A2* and *PDGFB* analysis by targeted
93 techniques (Sanger sequencing and QMPSF of all coding exons (12, 13)) followed by exome sequencing
94 in case of negative targeted screening, or by first-tier exome sequencing. From exome sequencing
95 data, rare variants in coding exons of known PFBC genes and other genes associated with brain
96 calcification were interpreted (22). In this retrospective analysis, we included a total of 113 patients
97 without any likely pathogenic or pathogenic variant in genes associated with brain calcification
98 (including heterozygous pathogenic variants in recessive genes).

99 Exomes were captured using Agilent Human all exons Surelect kits, V5 (n=11), V5+UTR (n=12) or
100 V6+UTR (n=90), and sequenced using an Illumina HiSeq4000 or Novaseq6000 sequencer with an
101 average depth of coverage target >100x, at the CNRGH sequencing center (Evry, France).
102 Bioinformatics pipelines were described in our previous reports (1, 22). Reference genome used in this
103 study is GRCh37.

104 This study was approved by the CERDE ethics committee of the Rouen University Hospital (E2023-40)
105 have therefore been performed in accordance with the ethical standards laid down in the 1964
106 Declaration of Helsinki and its later amendments.

107

108 *Coverage statistics*

109 We assessed whether 5'UTRs of all 6 PFBC genes were sufficiently covered in our exome sequencing
110 data. For that purpose, we retrieved the genome coordinates of the following Refseq transcripts

111 (GRCh37 human genome reference from the Alamut Visual software) and assessed the percentage of
112 bases covered by at least 10 or 20 reads: *SLC20A2*(NM_006749.5): chr8:42,397,097-42,396,964 and
113 42,330,172-42,329,909; *PDGFB*(NM_002608.4):chr22:39,640,987-39,639,969;
114 *PDFGRB*(NM_002609.4):chr5:149,534,958-149,535,403 and 149,516,616-149,516,611;
115 *XPR1*(NM_004736.4):chr1:180,601,158-180,601,337; *JAM2*(NM_021219.4):chr21:27,011,570-
116 27,012,133; *MYORG*(NM_020702.5modified):chr9 :34,376,896-34,376,791 and 34,373,004-
117 34,372,840. NB the latter transcript was used as the one corresponding to Ensembl 00000297625 with
118 natural initiation at position chr9:34,372,839 (after a common polymorphic nonsense variant in the
119 original NM_020702.5 reference). In Table S1 are displayed the statistics per capture kit, i.e. individual-
120 level percentage of bases covered at least 10x or 20x in each 5'UTR region of interest.

121

122 *Variant annotation and interpretation*

123 VCFs of all selected patients were reannotated using the 5utr ['suter'] (<https://github.com/leklab/5utr>)
124 and UTR annotator tools (23). We selected variants with a frequency below 0.001 in the gnomAD v2.1
125 database and a non-neutral effect based on at least one of these tools.

126 Then, we further assessed the putative consequences of selected variants using the following web
127 tools: NetStart (24) and ATGpr (25).

128 Segregation of selected variants in affected relatives was assessed, when DNA was available, using
129 either already available exome data or Sanger sequencing.

130 Variants were submitted to Clinvar (accession numbers: SUB14219653 SUB14219675)

131

132 *Reporter assay*

133 We set up a reporter assay similarly to our previous report on the *NIPBL* gene (26), with some
134 adaptations. Wild-type 5'UTR sequences (according to RefSeq data) of *PDGFB* (NM_002608.4:-1:-1019)
135 were synthesized and subcloned upstream of the GFP into the EcoRI/NotI sites of the pcDNA3.1(+)-C-
136 GFP vector (Genscript, Leiden, Netherlands). The Δ NotI deletion was generated using the Q5 Site-
137 Directed Mutagenesis System (E0554) V.2 Mutagenesis (New England Biolabs, Ipswich, Massachusetts,
138 USA) on the 5'UTR *PDGFB*^{wt}_pcDNA3.1(+)-C-GFP, according to manufacturer's instructions. The
139 sequences of the mutagenic oligonucleotides are: 5'-ATGAGCAAGGGCGAGGAGCTGTTCACC-3' and 5'-
140 GCCGACTCCGGGCCCGGC-3'. The c.-373C>G or c.-318C>T variants were introduced by site-directed
141 mutagenesis using the QuikChange II XL Site-Directed Mutagenesis Kit from Agilent (Santa Clara, CA,

142 USA) according to the manufacturer's instructions. The sequences of the mutagenic oligonucleotides
143 are: c.-373C>G F: 5'-ACACGGCAGTCCATGGTTCGTCTTCACTCG-3', c.-373C>G R: 5'-
144 CGAGTGAAGACGAACCATGGACTGCCGTGT-3', c.-318C>T F: 5'-GGGCGCCCCACATGGCTAGACGCCTCG-
145 3' and c.-318C>T R: 5'-CGAGGCGTCTAGCCATGTGGGGGCGCCC-3'. Plasmids containing GFP without
146 cloned 5'UTR, or with the 5'UTR of *PDGFB* in a wild-type context or with the c.-373C>G or c.-318C>T
147 variants were transiently transfected into HEK293 cells. HEK293 cells were grown in a culture medium
148 containing a 1:1 mix of DMEM and F12 (Gibco, Thermo Fischer Scientific, Waltham, MA, USA),
149 supplemented with 10% FCS (Eurobio, Les Ulis, France). For transfection, cells were grown in 12-well
150 plates and transfected with lipofectamine 3000 (Invitrogen, Thermo Fischer Scientific) according to the
151 manufacturer's instructions. Extraction was performed 48h later.

152

153 *RNA transcript level assessments*

154 Total RNA was extracted using the Nucleospin® RNA isolation kit (Macherey-Nagel, Düren, Germany),
155 according to the manufacturer's instructions. RNA was quantified by spectrophotometry (Nanodrop,
156 Thermo scientific). Reverse transcription was performed on 100ng RNA, using the Verso cDNA kit with
157 a mix (3:1) of random hexamers and oligodT primers (ThermoScientific). Gene expression was analysed
158 by digital droplet PCR (ddPCR) on a QX200 platform (Bio-Rad Laboratories, Hercules, CA, USA).
159 Relative GFP and neomycin-resistance gene expression in HEK293 cells was assessed by relative
160 quantification with RPL27, used as reference gene. The ddPCR protocol was performed as previously
161 described (22) using the same primers as in ref. (26).

162

163 *Protein extraction and immunoblot analysis*

164 Cells were homogenized in RIPA buffer (Pierce®, Thermo Fisher Scientific), supplemented with a
165 cocktail of protease inhibitors (Sigma-Aldrich, Saint-Louis, MI, USA) and phosphatase inhibitors (Halt
166 phosphatase, Thermo Fisher Scientific), as previously described (26). After 10 min on ice, lysates were
167 centrifuged (12,000 × g, 10 min, 4°C) and the supernatant containing soluble proteins was collected.
168 Protein concentration was measured using the DC Protein Assay Kit (Bio-Rad Laboratories, Hercules,
169 CA, USA). The GFP protein was resolved by 12% TGX Stain-Free gels (Bio-Rad Laboratories) and then
170 blotted onto nitrocellulose membranes using the Trans-Blot Turbo system (Bio-Rad Laboratories).
171 Membranes were immunoblotted with the monoclonal mouse anti-GFP antibody (1:1000)
172 (ref#11814460001, Roche, Bâle, Switzerland). After incubation with secondary peroxidase-labelled
173 anti-mouse antibody (1:10,000) from Jackson Immunoresearch Laboratories (WestGrove, PA, USA),

174 signals were detected with chemiluminescence reagents (ECL Clarity, Bio-Rad Laboratories). Signals
175 were acquired with a GBOX (Syngene, Cambridge, UK) monitored by the Gene Snap (Syngene)
176 software. The signal intensity in each lane was quantified using the Genetools software (Syngene) and
177 normalized with the Stain-Free signal quantified in the corresponding lane (ImageLab™ software, Bio-
178 Rad Laboratories).

179

180 *Statistics*

181 Normalized RNA and protein GFP levels were compared using ANOVA with a Dunnett's posttest. The
182 wild-type genotype was used as reference. Adjusted p-value: *** p =0.0002 ; **** p<0.0001.

183

184 **RESULTS**

185 *Identification of two upstream AUG-introducing variants*

186 We first checked whether exome sequencing data was of sufficient quality in the 5'UTR of the genes
187 of interest. Over time, we used three different capture kits for exome sequencing, the latter two being
188 specifically designed to cover UTR regions. Overall, all regions of interest were sufficiently covered
189 using the latest kit (V6+UTR, Agilent all exons human exome kit, used in the vast majority samples,
190 N=90). The least covered region was the 5'UTR of *PDGFB* albeit with an average percentage of bases
191 covered at least 10x of 92.1% across all 90 samples with the minimal value being 79.4% of bases in one
192 individual. The V5+UTR kit (n=12 samples) mainly missed exon 1 of the selected transcripts of *PDGFRB*,
193 *SLC20A2* and *MYORG*, while the V5 kit (n=12 samples) missed the same regions and with poor coverage
194 of the 5'UTR part of exon 2 of *SLC20A2* as well as the 5'UTR of *JAM2* and *PDGFB*.

195 After reannotation of 113 exomes of genetically unexplained unrelated PFBC probands, only the
196 *PDGFB* gene contained 5'UTR variants putatively affecting natural uORFs or creating novel uORFs. UTR
197 annotator and 5UTR annotation tools both detected two heterozygous variants introducing an
198 upstream AUG, while there was no such variant in the gnomAD v2.1 database: one with a predicted
199 high strength and an upstream stop codon 258 bp (86 codons) downstream
200 (NC_000022.10(NM_002608.4):c.-373C>G (hereafter named c.-373C>G) leading to a putative uORF,
201 and one with a predicted moderate strength and overlapping frameshift ORF (out of frame, oORF),
202 NC_000022.10(NM_002608.4):c.-318C>T (hereafter named c.-318C>T) (Figure 1). In addition, the UTR
203 annotator tool detected a NC_000022.10(NM_002608.4):c.-122_-115del variant in *PDGFB*, which was
204 predicted to remove a predicted natural upstream stop codon. As this variant was not predicted to

205 lead to haploinsufficiency, it was not further prioritized for Sanger sequencing confirmation and *in vitro*
206 assays.

207 Further *in silico* analysis of c.-373C>G and c.-318C>T variants showed that they were both detected as
208 a putative translation initiation site with higher scores than the natural site byATGpr and Netstart web
209 tools (supplementary Figure). Of note, the Netstart score of the natural site was low in wild-type
210 conditions and not considered as sufficient to initiate translation, contrary to the mutant condition.

211

212 *Clinical reports*

213 Patient EXT-986-002, carrying the c.-373C>G variant, is the proband of a family suggestive of autosomal
214 dominant PFBC (EXT-986). Segregation analysis in the affected father (EXT-986-001) showed that he
215 was a carrier of the same variant while the unaffected sister was not (EXT-986-003, normal CT scan at
216 the age of 50 years) (Figure 2).

217 The proband complained of memory difficulties at the age of 38 years but neuropsychological
218 examination was normal, as well as neurological examination. Given her known family history of brain
219 calcifications, she underwent a CT scan which showed bilateral calcifications of lenticular nuclei and
220 faint left cerebellar hemisphere calcifications (TCS=8/80, at age 46). She presented normal parathyroid
221 hormone, calcium and phosphorus blood levels.

222 Her father presented memory impairment from the age of 60 years as well as irritability and attention
223 deficit. Neurological examination at 61 years showed mild hypertonia of the right upper limb.
224 Neuropsychological assessment was indicative of dysexecutive and working memory impairment.
225 Brain imaging revealed calcifications of both lenticular nuclei and cerebellar hemispheres (TCS=10/80,
226 at age 61). He presented normal parathyroid hormone, calcium and phosphorus blood levels. His own
227 father (proband's paternal grandfather) had a history of bradykinesia, gait impairment and freezing
228 episodes with incontinence and swallowing problems. A paternal aunt of the proband's father also
229 showed gait disorder, while another aunt presented dementia.

230

231 Patient EXT-2263-001, carrying the c.-318C>T, is an apparently sporadic case, as no relative is known
232 to carry brain calcifications. While there was no brain imaging performed in relatives, to our
233 knowledge, two sib pairs and his mother were said to present cognitive impairment, one brother
234 committed suicide and another brother presented epilepsy. Brain calcifications were identified in the

235 proband in the context of acute vertigo at the age of 59 years, leading to a brain CT scan that revealed
236 bilateral calcifications of both lenticular nuclei and cerebellar hemispheres (TCS=13/80).

237

238 *Variants c.-373C>G and c.-318C>T in the 5'UTR of PDGFB affect protein levels in a reporter assay*

239 To assess the putative consequences of variants c.-373C>G and c.-318C>T on protein levels, we set up
240 an *in vitro* reporter assay similarly to a previous study by our group (26). HEK293 cells were transiently
241 transfected by a pcDNA3 plasmid containing the GFP coding sequence, under the control of the 5'UTR
242 of *PDGFB* in a wild-type context or with the variants (Figure 1). GFP protein levels were then assessed
243 by western blotting (WB). The normalized GFP expression was significantly decreased in c.-373C>G and
244 c.-318C>T variants compared to WT (Figure 3A). To determine if drops in protein levels were associated
245 with decreased mRNA levels, GFP mRNA levels were assessed by RT-ddPCR. Both variants led to a slight
246 but significant decrease in GFP mRNA levels, (Figure 3B), suggesting that the impact of the c.-373C>G
247 and c.-318C>T variants on GFP protein expression is mainly due to an impact on GFP translation,
248 accompanied by a small but measurable effect on mRNA steady-state levels.

249 As variant c.-318C>T was predicted to lead to a frameshift overlapping upstream ORF (our-of-frame
250 ORF, oORF), we sought to generate a slightly elongated mutant plasmid allowing the translation of a
251 longer GFP-tagged protein from the newly created upstream start codon. To this aim, we used plasmids
252 containing the 8bp Not1 restriction site located between the 5'UTR of *PDGFB* and the translation start
253 of the GFP (Figure 1). This insertion re-establishes the reading frame between the upstream start
254 codon and the GFP cDNA. WB analyses showed the presence of a protein at the expected 40-kDa
255 weight (Figure 3C), indicating a translation from the novel AUG created by the c.-318C>T variant and
256 resulting in the predicted elongated in-frame ORF (eORF).

257 These experiments suggest that both variants are functional. Interpreted in the context of the *PDGFB*
258 gene, these variants are thus predicted to lead to reduced translation of PDGFB in the c.-373C>G
259 variant carrier and to the production of a frameshift protein instead of PDGFB, at least partially, in the
260 c.-318C>T variant carrier. Taken together, these results suggest that these variants are likely
261 pathogenic (class 4) following the American College of Medical Genetics and Genomics and the
262 Association for Molecular Pathology (ACMG-AMP) recommendations (27), using the following
263 arguments: PS3 (well established *in vitro* assessment supportive of a damaging effect on the gene or
264 gene product) PM2 (absent from controls), PP4 (patient's phenotype) and PP1 (cosegregation), for
265 family EXT-0986.

266

267 **DISCUSSION**

268 More than 50% of the patients diagnosed with PFBC remain genetically unexplained. While it is very
269 likely that all PFBC-causal genes have not been identified yet, variants in non-coding regions of known
270 genes still represent a source of missed diagnoses. However, such non-coding variants remain difficult
271 to interpret, despite recent publication of recommendations (28) and in the context of increasing
272 access to genome sequencing. In PFBC, examples of pathogenic non-coding variants are scarce. We
273 previously reported a non-coding deletion of an enhancer of *SLC20A2* transcription in a family with
274 autosomal dominant PFBC (22). In addition, a deletion of the 5'UTR region and the promoter of
275 *SLC20A2* as well as other upstream genes was identified in a Finnish patient (29). Most non-coding
276 variants require additional experiments to conclude on pathogenicity. While the above-mentioned
277 deletion of the 5'UTR of *SLC20A2* was large enough to speculate on a haploinsufficient effect without
278 any additional experiment, in absence of any promoter (29), mRNA quantification and *in vitro*
279 modelling of the *SLC20A2* enhancer deletion was required to conclude on variant pathogenicity (22).
280 Here, we hypothesized that variants in the 5'UTR of known genes may create upstream open reading
281 frames and lead to a loss of function, thus representing another source of non-coding pathogenic
282 variants.

283 Interestingly, only the *PDGFB* gene was affected by upstream AUG-introducing variants, and we could
284 not detect any such variant in *SLC20A2* despite good sequencing coverage in most patients, while this
285 gene represents the main PFBC gene in terms of relative frequency of pathogenic variants in PFBC (30).
286 *SLC20A2* main transcripts NM_006749.5 and NM_001257180.2 contain 398 and 465 bp-long 5'UTRs.
287 *PDGFB* NM_002608.4 transcript, corresponding to the second most frequently mutated gene in PFBC
288 patients, is 1019 bp-long. Both genes have predicted natural uORFs, suggesting that such a mechanism
289 may be used naturally to regulate both genes' expression. Conversely, it is less surprising not to identify
290 upstream AUG-introducing variants in *PDGFRB* and *XPR1*, as there are currently no purely
291 haploinsufficient variants identified in the coding sequence of these genes. The recently identified
292 autosomal recessive gene *JAM2* seems to be an extremely rare cause of PFBC with no pathogenic
293 variant in our series. The other recessive gene, *MYORG*, has a relatively small 5'UTR (227 bp based on
294 the corrected ENST00000297625.8 transcript). Interestingly, *MYORG* is often annotated using the
295 NM_020702.5 transcript, with a natural start codon 34 codons upstream of the start codon in the
296 ENST00000297625.8 transcript. However, there is a very common nonsense polymorphism
297 (rs4879782) at position p.23 in the NM_020702.5 sequence, which could represent some kind of uORF
298 in the variant carriers, with no known pathogenic consequence, as 6.7% of individuals in the gnomAD
299 v3 database are homozygous carriers. This does not make this gene a good candidate for a strong
300 regulation through uAUG-introducing variants.

301 Overall, two variants creating an upstream ORF were identified here (1.7%), in one of the largest series
302 of PFBC negative exomes of unrelated probands. Such a mechanism is probably a rare mutational
303 event, similarly to other diseases such as Cornelia de Lange syndrome (26) or other developmental
304 disorders (31). As reannotation of existing variant files from exome data produces a manageable
305 number of variants per individual, it appears reasonable to perform a reannotation and
306 reinterpretation of negative exomes before proposing novel sequencing techniques, as whole genome
307 sequencing, even if the enrichment method was not optimized for UTRs.

308 However, and as for other non-coding variants, it remains necessary to assess the consequences of
309 variants *in vitro* after they have been prioritized by specific annotation tools. Indeed, the
310 bioinformatics predictions are not sufficient to provide evidence of an actual effect, and segregation
311 data often remain too limited. Here, we relied on a reporter assay with mRNA and protein detection
312 as a readout. There are some limitations to such assays. First, they do not use the natural translation
313 initiation site of the protein of interest, but the one of the reporter protein. Here, the measured effect
314 of the upstream AUGs was strong in both mutant contexts, as compared to the translation initiation
315 site of GFP. We hypothesize that it might be strong in a *PDGFB* context as well. However, for some
316 other genes, the sensitivity of the assay may be different. Second, this assay is based on the expression
317 of vectors in cells and not necessarily at similar levels as in the patients' cells of interest. Despite this,
318 we previously found a good correlation between the results of our reporter assay assessing the
319 consequences of variants in the 5'UTR of *NIPBL* and protein levels in patient-derived lymphoblastoid
320 cell lines (26). Unfortunately, we could not assess the consequences of the variant in patients' tissue
321 here. Third, the activity of the protein is not measured by our assay. However, drastic reduction of
322 protein levels in both contexts is compatible with the mechanism associated with most *PDGFB*
323 pathogenic variants, i.e. haploinsufficiency, so that an activity assay is not required in this context.

324 Both patients presented PFBC with some radiological and clinical diversity. We cannot exclude that the
325 alleles detected here have a non-complete effect, as translation from the natural initiation site may
326 still exist. However, the phenotypic expression of PFBC is known to be very diverse within and between
327 families, so that no specific interpretation on the phenotypic expression or clinical or radiological
328 severity can be drawn from these families.

329 In conclusion, we identified a source of non-coding likely pathogenic variants in genetically
330 unexplained PFBC patients previously assessed by exome sequencing. Rare variants in the 5'UTR of the
331 *PDGFB* gene should be interpreted in a clinical setting, not to miss putatively pathogenic variants in
332 PFBC patients. It remains unclear whether variants in the 5'UTR of the other PFBC genes could
333 represent a source of pathogenic variants.

334

335 **Data Availability Statement**

336 All relevant data is available in the manuscript or supplement

337

338 **Acknowledgements**

339 We are grateful to the patients and their families.

340

341 **Funding**

342 We thank SFN and JNLF for A.B. research fellowship. This study was supported by grants from the
343 French National Research Agency (CALCIPHOS, ANR-17-CE14-0008 to GN) and from Conseil Régional
344 de Haute Normandie—APEREC 2014 no. 2014-19 in the context of Appel d’Offres Jeunes Chercheurs
345 (CHU de Rouen to GN). This study was cosupported by European Union and Région Normandie, more
346 specifically in the context of the Recherche Innovation Normandie (RIN 2018 to GN). Europe gets
347 involved in Normandie with the European Regional Development Fund. The CEA-CNRGH sequencing
348 platform was supported by the France Génomique National infrastructure, funded as part of the «
349 Investissements d’Avenir » program managed by the Agence Nationale pour la Recherche (contract
350 ANR-10-INBS-09)

351

352 **Author Contribution Statement**

353 Study design and study supervision: GN, ML

354 Data acquisition and data analysis: all authors

355 Drafting the manuscript: GN, ARL, ML

356

357 **Ethical Approval**

358 This study was approved by the CERDE ethics committee of the Rouen University Hospital (E2023-40)
359 have therefore been performed in accordance with the ethical standards laid down in the 1964
360 Declaration of Helsinki and its later amendments.

361

362 **Competing interests**

363 Nothing to report.

364 **FIGURE LEGENDS**

365 **Figure 1. Schematic representation of the identified PDGFB c.-373C>G and c.-318C>T variants and**
366 **molecular constructs**

367 (A) PDGFB wild-type and c.-373C>G and c.-318C>T cDNA sequences. Arrows indicate the main Open
368 Reading Frame (mORF, "+1") and newly created upstream (non-overlapping) ORF (uORF) and out-of
369 frame overlapping ORF (oORF) initiation codons. (B) Molecular constructs used in the *in vitro* GFP
370 reporter assay to investigate the effect of the variants.

371
372
373

374 **Figure 2. Reduced pedigrees of Family EXT-986 (left panel) and EXT-2263 (right panel).**

375 Black-filled symbols indicate patients with brain calcifications. Unless otherwise specified, white
376 symbols indicate individuals without any brain CT available. Arrows indicate probands.
377 Stars indicate individuals with available DNA

378

379 **Figure 3. Functional impact of the PDGFB c.-373C>G and c.-318C>T variants on GFP expression using**
380 **an *in vitro* GFP reporter assay.**

381 A. Left panel, western blot showing the expression of GFP protein (Black arrow head) under the
382 control of the 5'UTR of PDGFB, in a wild-type (wt) condition (n=2), or carrying the c.-373G>C or c.-
383 318C>T variants (n=3). Total protein was used as loading control with Stain-free (SF) technology. A
384 plasmid containing the GFP coding sequence without the PDGFB 5'UTR (GFP) and an empty vector
385 (mock) were used as positive and negative controls, respectively. Normalized expression of the GFP
386 protein reported to the SF staining is reported in the graph (Right panel). Bar graph: mean \pm standard
387 deviation of relative GFP protein levels. WT mean was arbitrarily set at 1.

388 B. Normalized GFP mRNA levels reported to the neomycin resistance gene (wt n=4, variants n=6). Bar
389 graph: mean \pm standard deviation of relative GFP mRNA levels. WT mean was arbitrarily set at 1.

390 C. Western blot showing the expression of GFP under the control of the 5'UTR of PDGFB when the
391 reading frame is re-established between the upstream start codon resulting from the c.-318C>T
392 variant and the GFP cDNA (+ NotI), in a wild-type (wt) condition or with the c.-318C>T variant (n=2).
393 We detected the presence of a larger mutant protein (gray arrowhead) with the elongated mutant,
394 probably resulting from the use of the new AUG.

395

396

397

398 **Supplement**
399 **Supplementary figure.** In silico predictions
400 **Supplementary table.** Coverage statistics
401
402

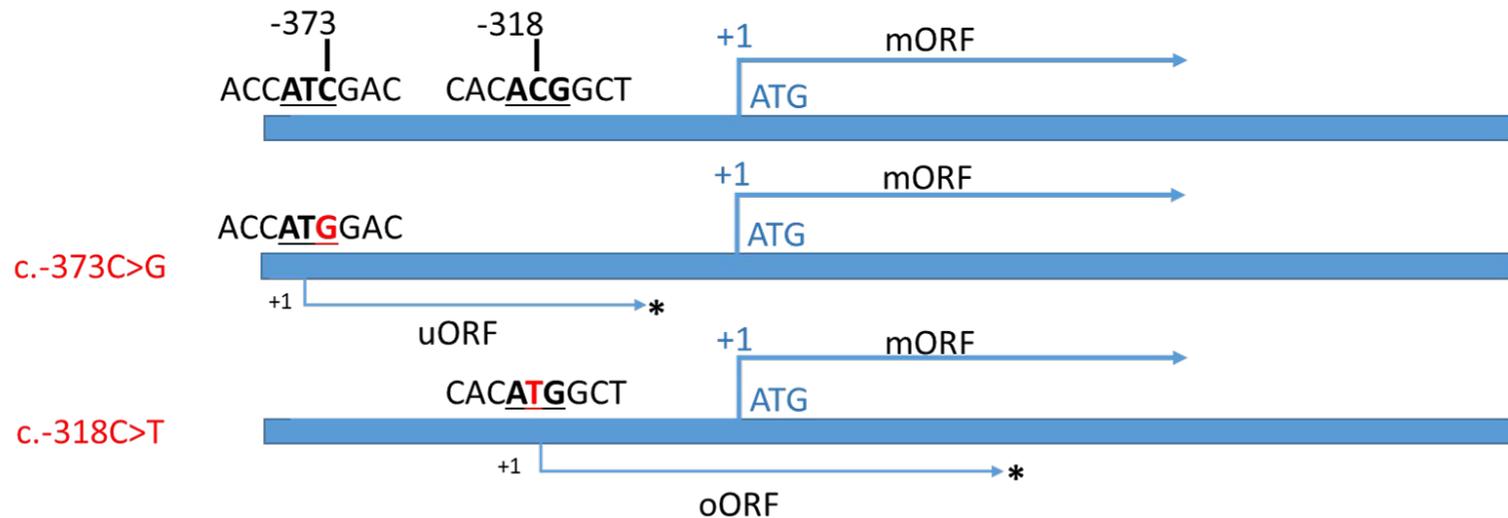
403 **REFERENCES**

- 404 1. Grangeon L, Wallon D, Charbonnier C, Quenez O, Richard AC, Rousseau S, et al. Biallelic MYORG
405 mutation carriers exhibit primary brain calcification with a distinct phenotype. *Brain : a journal of*
406 *neurology*. 2019;142(6):1573-86.
- 407 2. Balck A, Schaake S, Kuhnke NS, Domingo A, Madoev H, Margolesky J, et al. Genotype-
408 Phenotype Relations in Primary Familial Brain Calcification: Systematic MDSGene Review. *Movement*
409 *disorders : official journal of the Movement Disorder Society*. 2021;36(11):2468-80.
- 410 3. Wang C, Li Y, Shi L, Ren J, Patti M, Wang T, et al. Mutations in SLC20A2 link familial idiopathic
411 basal ganglia calcification with phosphate homeostasis. *Nature genetics*. 2012;44(3):254-6.
- 412 4. Nicolas G, Pottier C, Maltete D, Coutant S, Rovelet-Lecrux A, Legallic S, et al. Mutation of the
413 PDGFRB gene as a cause of idiopathic basal ganglia calcification. *Neurology*. 2013;80(2):181-7.
- 414 5. Keller A, Westenberger A, Sobrido MJ, Garcia-Murias M, Domingo A, Sears RL, et al. Mutations
415 in the gene encoding PDGF-B cause brain calcifications in humans and mice. *Nature genetics*.
416 2013;45(9):1077-82.
- 417 6. Legati A, Giovannini D, Nicolas G, Lopez-Sanchez U, Quintans B, Oliveira JR, et al. Mutations in
418 XPR1 cause primary familial brain calcification associated with altered phosphate export. *Nature*
419 *genetics*. 2015;47(6):579-81.
- 420 7. Yao XP, Cheng X, Wang C, Zhao M, Guo XX, Su HZ, et al. Biallelic Mutations in MYORG Cause
421 Autosomal Recessive Primary Familial Brain Calcification. *Neuron*. 2018;98(6):1116-23 e5.
- 422 8. Cen Z, Chen Y, Chen S, Wang H, Yang D, Zhang H, et al. Biallelic loss-of-function mutations in
423 JAM2 cause primary familial brain calcification. *Brain : a journal of neurology*. 2020;143(2):491-502.
- 424 9. Schottlaender LV, Abeti R, Jaunmuktane Z, Macmillan C, Chelban V, O'Callaghan B, et al. Bi-
425 allelic JAM2 Variants Lead to Early-Onset Recessive Primary Familial Brain Calcification. *American*
426 *journal of human genetics*. 2020;106(3):412-21.
- 427 10. Ramos EM, Oliveira J, Sobrido MJ, Coppola G. Primary Familial Brain Calcification. In: Adam
428 MP, Mirzaa GM, Pagon RA, Wallace SE, Bean LJH, Gripp KW, et al., editors. *GeneReviews((R))*. Seattle
429 (WA)1993.
- 430 11. Zhao M, Su HZ, Zeng YH, Sun Y, Guo XX, Li YL, et al. Loss of function of CMPK2 causes
431 mitochondria deficiency and brain calcification. *Cell discovery*. 2022;8(1):128.
- 432 12. Nicolas G, Rovelet-Lecrux A, Pottier C, Martinaud O, Wallon D, Vernier L, et al. PDGFB partial
433 deletion: a new, rare mechanism causing brain calcification with leukoencephalopathy. *Journal of*
434 *molecular neuroscience : MN*. 2014;53(2):171-5.
- 435 13. David S, Ferreira J, Quenez O, Rovelet-Lecrux A, Richard AC, Verin M, et al. Identification of
436 partial SLC20A2 deletions in primary brain calcification using whole-exome sequencing. *European*
437 *journal of human genetics : EJHG*. 2016;24(11):1630-4.
- 438 14. Anheim M, Lopez-Sanchez U, Giovannini D, Richard AC, Touhami J, N'Guyen L, et al. XPR1
439 mutations are a rare cause of primary familial brain calcification. *Journal of neurology*.
440 2016;263(8):1559-64.
- 441 15. Lopez-Sanchez U, Nicolas G, Richard AC, Maltete D, Charif M, Aygnac X, et al. Characterization
442 of XPR1/SLC53A1 variants located outside of the SPX domain in patients with primary familial brain
443 calcification. *Scientific reports*. 2019;9(1):6776.
- 444 16. Lenglez S, Sablon A, Fenelon G, Boland A, Deleuze JF, Boutoleau-Brettonniere C, et al. Distinct
445 functional classes of PDGFRB pathogenic variants in primary familial brain calcification. *Human*
446 *molecular genetics*. 2022;31(3):399-409.
- 447 17. Vanlandewijck M, Lebouvier T, Andaloussi Mae M, Nahar K, Hornemann S, Kenkel D, et al.
448 Functional Characterization of Germline Mutations in PDGFB and PDGFRB in Primary Familial Brain
449 Calcification. *PLoS one*. 2015;10(11):e0143407.
- 450 18. Barbosa C, Peixeiro I, Romao L. Gene expression regulation by upstream open reading frames
451 and human disease. *PLoS genetics*. 2013;9(8):e1003529.

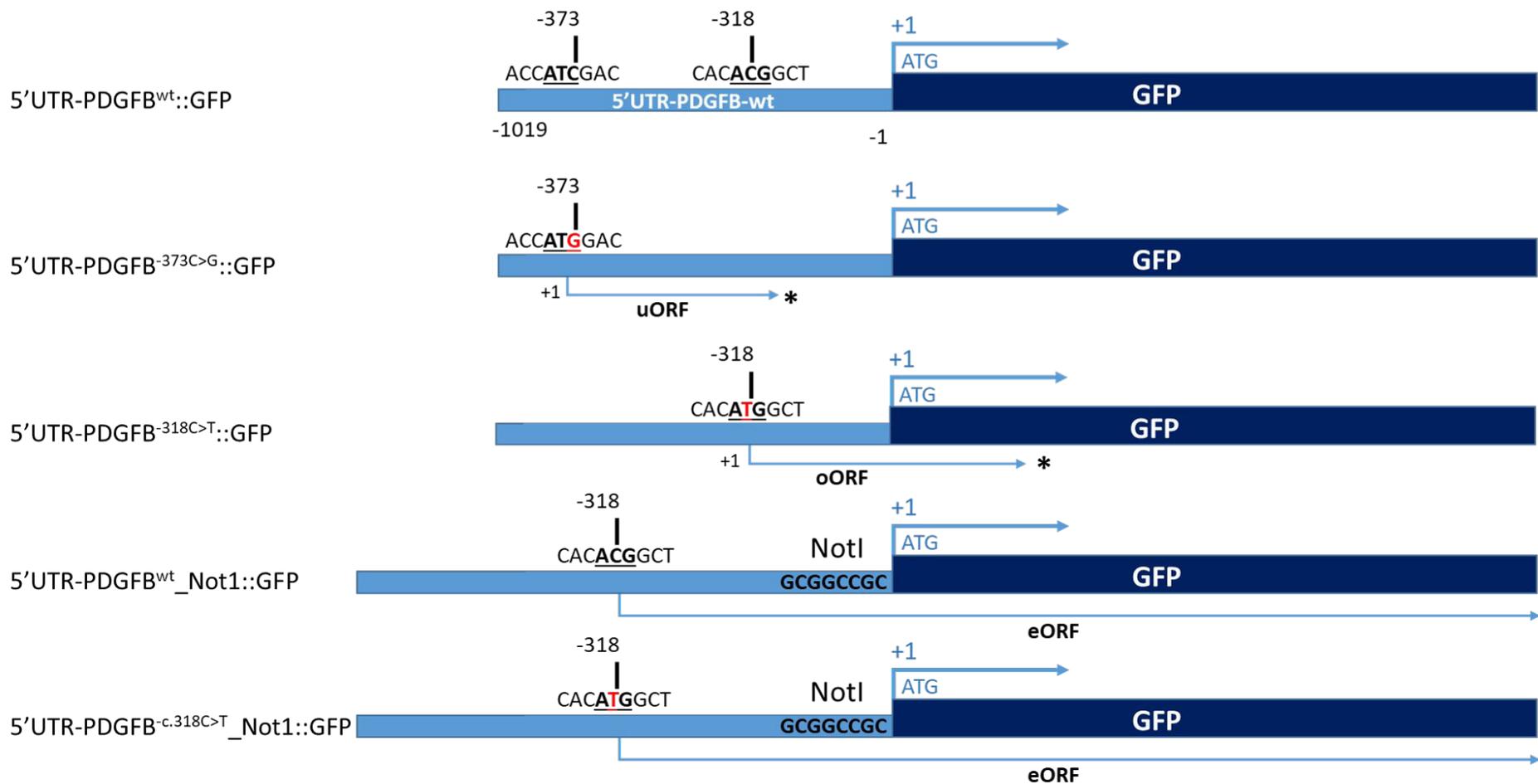
- 452 19. Whiffin N, Karczewski KJ, Zhang X, Chothani S, Smith MJ, Evans DG, et al. Characterising the
453 loss-of-function impact of 5' untranslated region variants in 15,708 individuals. *Nature*
454 *communications*. 2020;11(1):2523.
- 455 20. Kute PM, Soukarieh O, Tjeldnes H, Tregouet DA, Valen E. Small Open Reading Frames, How to
456 Find Them and Determine Their Function. *Frontiers in genetics*. 2021;12:796060.
- 457 21. Nicolas G, Pottier C, Charbonnier C, Guyant-Marechal L, Le Ber I, Pariente J, et al. Phenotypic
458 spectrum of probable and genetically-confirmed idiopathic basal ganglia calcification. *Brain : a journal*
459 *of neurology*. 2013;136(Pt 11):3395-407.
- 460 22. Cassinari K, Rovelet-Lecrux A, Tury S, Quenez O, Richard AC, Charbonnier C, et al.
461 Haploinsufficiency of the Primary Familial Brain Calcification Gene SLC20A2 Mediated by Disruption of
462 a Regulatory Element. *Movement disorders : official journal of the Movement Disorder Society*.
463 2020;35(8):1336-45.
- 464 23. Zhang X, Wakeling M, Ware J, Whiffin N. Annotating high-impact 5'untranslated region variants
465 with the UTRannotator. *Bioinformatics*. 2021;37(8):1171-3.
- 466 24. Pedersen AG, Nielsen H. Neural network prediction of translation initiation sites in eukaryotes:
467 perspectives for EST and genome analysis. *Proceedings International Conference on Intelligent*
468 *Systems for Molecular Biology*. 1997;5:226-33.
- 469 25. Salamov AA, Nishikawa T, Swindells MB. Assessing protein coding region integrity in cDNA
470 sequencing projects. *Bioinformatics*. 1998;14(5):384-90.
- 471 26. Coursimault J, Rovelet-Lecrux A, Cassinari K, Brischoux-Boucher E, Saugier-veber P,
472 Goldenberg A, et al. uORF-introducing variants in the 5'UTR of the NIPBL gene as a cause of Cornelia
473 de Lange syndrome. *Human mutation*. 2022;43(9):1239-48.
- 474 27. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the
475 interpretation of sequence variants: a joint consensus recommendation of the American College of
476 Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine :*
477 *official journal of the American College of Medical Genetics*. 2015;17(5):405-24.
- 478 28. Ellingford JM, Ahn JW, Bagnall RD, Baralle D, Barton S, Campbell C, et al. Recommendations for
479 clinical interpretation of variants found in non-coding regions of the genome. *Genome medicine*.
480 2022;14(1):73.
- 481 29. Pasanen P, Makinen J, Myllykangas L, Guerreiro R, Bras J, Valori M, et al. Primary familial brain
482 calcification linked to deletion of 5' noncoding region of SLC20A2. *Acta neurologica Scandinavica*.
483 2017;136(1):59-63.
- 484 30. Ramos EM, Carecchio M, Lemos R, Ferreira J, Legati A, Sears RL, et al. Primary brain
485 calcification: an international study reporting novel variants and associated phenotypes. *European*
486 *journal of human genetics : EJHG*. 2018;26(10):1462-77.
- 487 31. Wright CF, Quaife NM, Ramos-Hernandez L, Danecek P, Ferla MP, Samocha KE, et al. Non-
488 coding region variants upstream of MEF2C cause severe developmental disorder through three distinct
489 loss-of-function mechanisms. *American journal of human genetics*. 2021;108(6):1083-94.

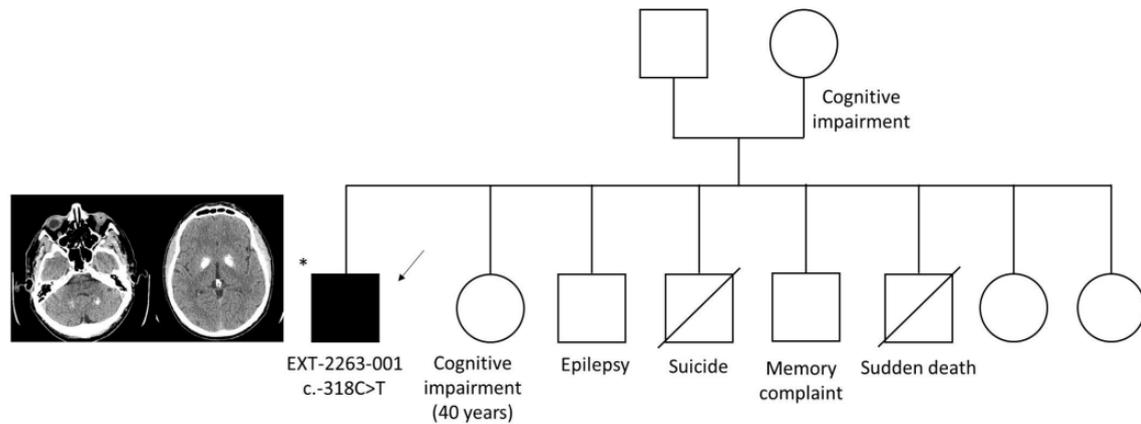
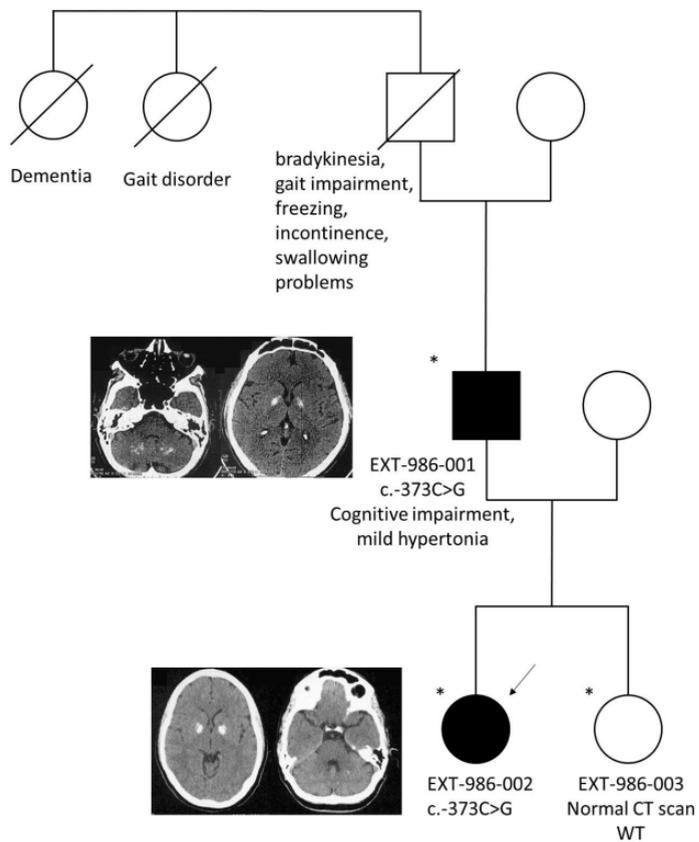
490

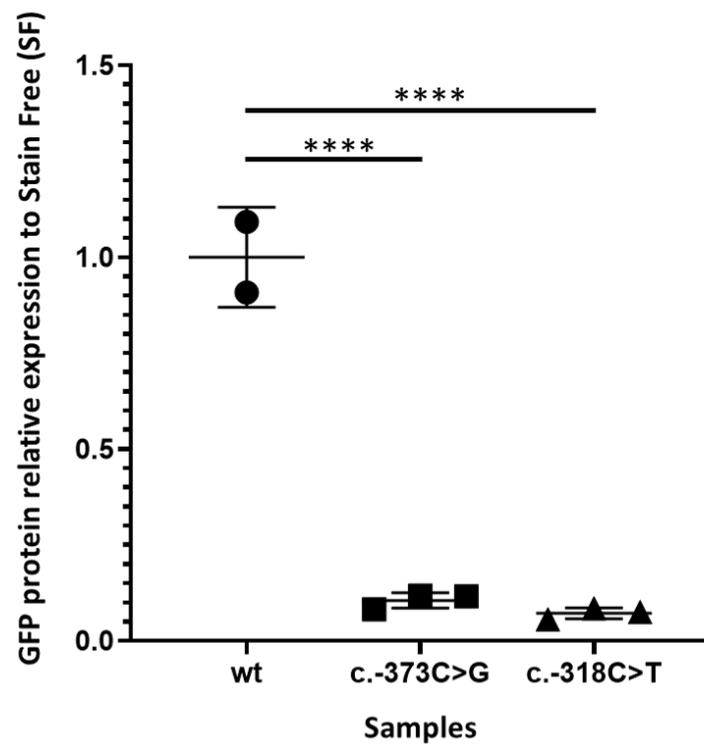
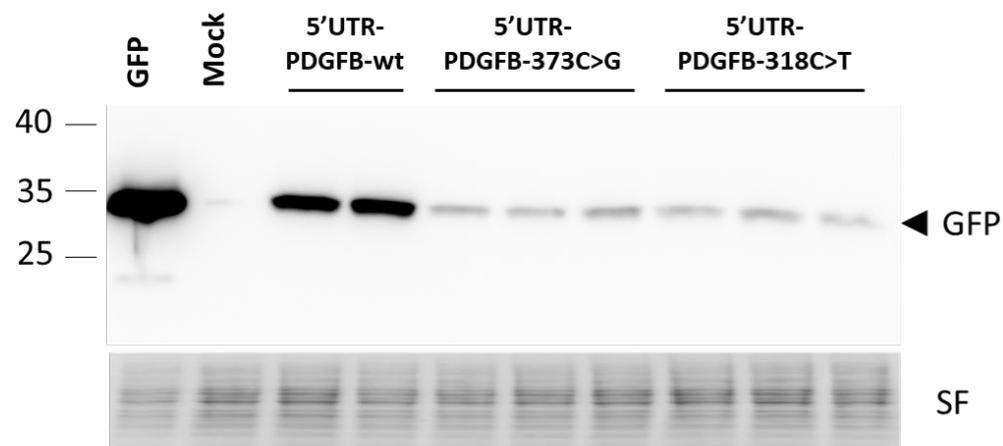
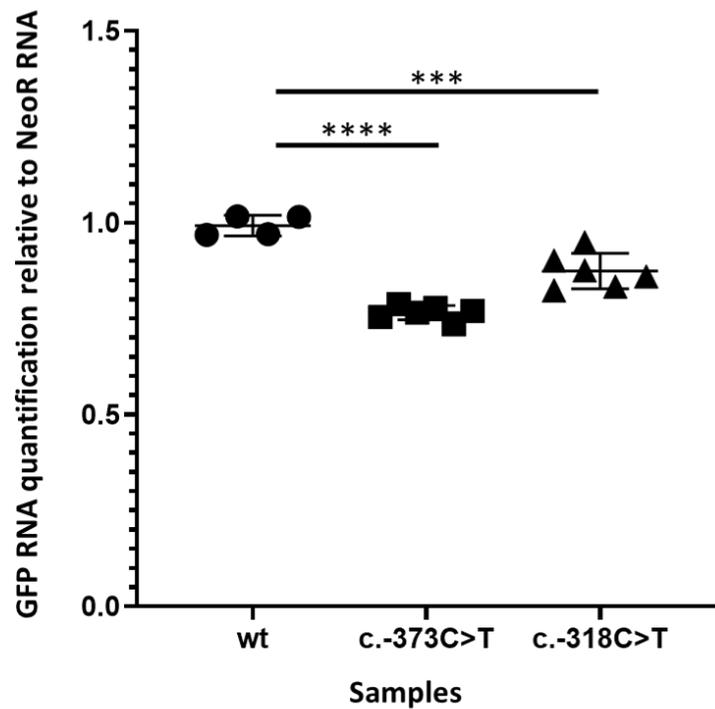
A



B





A**B****C**